

PCTWORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ : C12N 15/31, 15/52, 15/82, 15/70, 5/10, 1/21, C12P 7/64, A01H 5/00	A1	(11) International Publication Number: WO 98/55625 (43) International Publication Date: 10 December 1998 (10.12.98)
(21) International Application Number: PCT/US98/11639 (22) International Filing Date: 4 June 1998 (04.06.98) (30) Priority Data: 60/048,650 4 June 1997 (04.06.97) US (71) Applicant: CALGENE, LLC [US/US]; 1920 Fifth Street, Davis, CA 95616 (US). (72) Inventors: FACCIOTTI, Daniel; 2636 Lafayette Drive, Davis, CA 95616 (US). METZ, James, George; 2803 Belhaven Place, Davis, CA 95616 (US). LASSNER, Michael; 721 Falcon Avenue, Davis, CA 95616 (US). (74) Agent: RAE-VENTER, Barbara; Rae-Venter Law Group, P.C., P.O. Box 60039, Palo Alto, CA 94306 (US).		(81) Designated States: BR, CA, IL, JP, MX, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE). Published <i>With international search report. Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i>
(54) Title: PRODUCTION OF POLYUNSATURATED FATTY ACIDS BY EXPRESSION OF POLYKETIDE-LIKE SYNTHESIS GENES IN PLANTS (57) Abstract The present invention relates to compositions and methods for preparing polyunsaturated long chain fatty acids in plants, plant parts and plant cells, such as leaves, roots, fruits and seeds. Nucleic acid sequences and constructs encoding PKS-like genes required for the poly-unsaturated long chain fatty acid production, including the genes responsible for eicosapentenoic acid production of <i>Shewanella putrefaciens</i> and novel genes associated with the production of docosahexenoic acid in <i>Vibrio marinus</i> are used to generate transgenic plants, plant parts and cells which contain and express one or more transgenes encoding one or more of the PKS-like genes associated with such long chain polyunsaturated fatty acid production. Expression of the PKS-like genes in the plant system permits the large scale production of polyunsaturated long chain fatty acids such as eicosapentenoic acid and docosahexenoic acid for modification of the fatty acid profile of plants, plant parts and tissues. Manipulation of the fatty acid profiles allows for the production of commercial quantities of novel plant oils and products.		

BEST AVAILABLE COPY

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

PRODUCTION OF POLYUNSATURATED FATTY ACIDS BY EXPRESSION OF POLYKETIDE-LIKE SYNTHESIS GENES IN PLANTS

INTRODUCTION

Field of the Invention

This invention relates to modulating levels of enzymes and/or enzyme components capable of modifying long chain poly-unsaturated fatty acids (PUFAs) in a host cell, and constructs and methods for producing PUFAs in a host cell. The invention is exemplified by production of eicosapentenoic acid (EPA) using genes derived from *Shewanella putrefaciens* and *Vibrio marinus*.

Background

Two main families of poly-unsaturated fatty acids (PUFAs) are the ω 3 fatty acids, exemplified by eicosapentenoic acid, and the ω 6 fatty acids, exemplified by arachidonic acid. PUFAs are important components of the plasma membrane of the cell, where they can be found in such forms as phospholipids, and also can be found in triglycerides. PUFAs also serve as precursors to other molecules of importance in human beings and animals, including the prostacyclins, leukotrienes and prostaglandins. Long chain PUFAs of importance include docosahexenoic acid (DHA) and eicosapentenoic acid (EPA), which are found primarily in different types of fish oil, gamma-linolenic acid (GLA), which is found in the seeds of a number of plants, including evening primrose (*Oenothera biennis*), borage (*Borago officinalis*) and black currants (*Ribes nigrum*), stearidonic acid (SDA), which is found in marine oils and plant seeds, and arachidonic acid (ARA), which along with GLA is found in filamentous fungi. ARA can be purified from animal tissues including liver and adrenal gland. Several genera of marine bacteria are known which synthesize either EPA or DHA. DHA is present in human milk along with ARA.

PUFAs are necessary for proper development, particularly in the developing infant brain, and for tissue formation and repair. As an example, DHA, is an important constituent of many human cell membranes, in particular nervous cells (gray matter), muscle cells, and spermatozoa and believed to affect the development of brain functions in general and to be essential for the development of eyesight. EPA and DHA have a number of nutritional and pharmacological uses. As an example adults affected by diabetes (especially non insulin-dependent) show deficiencies and imbalances in their

levels of DHA which are believed to contribute to later coronary conditions. Therefore a diet balanced in DHA may be beneficial to diabetics.

For DHA, a number of sources exist for commercial production including a variety of marine organisms, oils obtained from cold water marine fish, and egg yolk fractions. The purification of DHA from fish sources is relatively expensive due to technical difficulties, making DHA expensive and in short supply. In algae such as *Amphidinium* and *Schyzochytrium* and marine fungi such as *Thraustochytrium* DHA may represent up to 48% of the fatty acid content of the cell. A few bacteria also are reported to produce DHA. These are generally deep sea bacteria such as *Vibrio marinus*. For ARA, microorganisms including the genera *Mortierella*, *Entomophthora*, *Phytium* and *Porphyridium* can be used for commercial production. Commercial sources of SDA include the genera *Trichodesma* and *Echium*. Commercial sources of GLA include evening primrose, black currants and borage. However, there are several disadvantages associated with commercial production of PUFAs from natural sources. Natural sources of PUFA, such as animals and plants, tend to have highly heterogeneous oil compositions. The oils obtained from these sources can require extensive purification to separate out one or more desired PUFA or to produce an oil which is enriched in one or more desired PUFA.

Natural sources also are subject to uncontrollable fluctuations in availability. Fish stocks may undergo natural variation or may be depleted by overfishing. Animal oils, and particularly fish oils, can accumulate environmental pollutants. Weather and disease can cause fluctuation in yields from both fish and plant sources. Cropland available for production of alternate oil-producing crops is subject to competition from the steady expansion of human populations and the associated increased need for food production on the remaining arable land. Crops which do produce PUFAs, such as borage, have not been adapted to commercial growth and may not perform well in monoculture. Growth of such crops is thus not economically competitive where more profitable and better established crops can be grown. Large -scale fermentation of organisms such as *Shewanella* also is expensive. Natural animal tissues contain low amounts of ARA and are difficult to process. Microorganisms such as *Porphyridium* and *Shewanella* are difficult to cultivate on a commercial scale.

Dietary supplements and pharmaceutical formulations containing PUFAs can retain the disadvantages of the PUFA source. Supplements such as fish oil capsules can

contain low levels of the particular desired component and thus require large dosages. High dosages result in ingestion of high levels of undesired components, including contaminants. Care must be taken in providing fatty acid supplements, as overaddition may result in suppression of endogenous biosynthetic pathways and lead to competition with other necessary fatty acids in various lipid fractions *in vivo*, leading to undesirable results. For example, Eskimos having a diet high in ω 3 fatty acids have an increased tendency to bleed (U.S. Pat. No. 4,874,603). Fish oils have unpleasant tastes and odors, which may be impossible to economically separate from the desired product, such as a food supplements. Unpleasant tastes and odors of the supplements can make such regimens involving the supplement undesirable and may inhibit compliance by the patient.

A number of enzymes have been identified as being involved in PUFA biosynthesis. Linoleic acid (LA, 18:2 Δ 9, 12) is produced from oleic acid (18:1 Δ 9) by a Δ 12-desaturase. GLA (18:3 Δ 6, 9, 12) is produced from linoleic acid (LA, 18:2 Δ 9, 12) by a Δ 6-desaturase. ARA (20:4 Δ 5, 8, 11, 14) is produced from DGLA (20:3 Δ 8, 11, 14), catalyzed by a Δ 5-desaturase. Eicosapentenoic acid (EPA) is a 20 carbon, omega 3 fatty acid containing 5 double bonds (Δ 5, 8, 11, 14, 17), all in the *cis* configuration. EPA, and the related DHA (Δ 4, 7, 10, 13, 16, 19, C22:6) are produced from oleic acid by a series of elongation and desaturation reactions. Additionally, an elongase (or elongases) is required to extend the 18 carbon PUFAs out to 20 and 22 carbon chain lengths. However, animals cannot convert oleic acid (18:1 Δ 9) into linoleic acid (18:2 Δ 9, 12). Likewise, μ -linolenic acid (ALA, 18:3 Δ 9, 12, 15) cannot be synthesized by mammals. Other eukaryotes, including fungi and plants, have enzymes which desaturate at positions Δ 12 and Δ 15. The major poly-unsaturated fatty acids of animals therefore are either derived from diet and/or from desaturation and elongation of linoleic acid (18:2 Δ 9, 12) or μ -linolenic acid (18:3 Δ 9, 12, 15).

Poly-unsaturated fatty acids are considered to be useful for nutritional, pharmaceutical, industrial, and other purposes. An expansive supply of poly-unsaturated fatty acids from natural sources and from chemical synthesis are not sufficient for commercial needs. Because a number of separate desaturase and elongase enzymes are required for fatty acid synthesis from linoleic acid (LA, 18:2 Δ 9, 12), common in most plant species, to the more saturated and longer chain PUFAs, engineering plant host cells for the expression of EPA and DHA may require expression of five or six separate

enzyme activities to achieve expression, at least for EPA and DHA, and for production of quantities of such PUFAs additional engineering efforts may be required, for instance the down regulation of enzymes competing for substrate, engineering of higher enzyme activities such as by mutagenesis or targeting of enzymes to plastid organelles. Therefore it is of interest to obtain genetic material involved in PUFA biosynthesis from species that naturally produce these fatty acids and to express the isolated material alone or in combination in a heterologous system which can be manipulated to allow production of commercial quantities of PUFAs.

Relevant Literature

Several genera of marine bacteria have been identified which synthesize either EPA or DHA (DeLong and Yayanos, *Applied and Environmental Microbiology* (1986) 51: 730-737). Researchers of the Sagami Chemical Research Institute have reported EPA production in *E. coli* which have been transformed with a gene cluster from the marine bacterium, *Shewanella putrefaciens*. A minimum of 5 open reading frames (ORFs) are required for fatty acid synthesis of EPA in *E. coli*. To date, extensive characterization of the functions of the proteins encoded by these genes has not been reported (Yazawa (1996) *Lipids* 31, S-297; WO 93/23545; WO 96/21735).

The protein sequence of open reading frame (ORF) 3 as published by Yazawa, USPN 5,683,898 is not a functional protein. Yazawa defines the protein as initiating at the methionine codon at nucleotides 9016-9014 of the *Shewanella* PKS-like cluster (Genbank accession U73935) and ending at the stop codon at nucleotides 8185-8183 of the *Shewanella* PKS-like cluster. However, when this ORF is expressed under control of a heterologous promoter in an *E. coli* strain containing the entire PKS-like cluster except ORF 3, the recombinant cells do not produce EPA.

Polyketides are secondary metabolites the synthesis of which involves a set of enzymatic reactions analogous to those of fatty acid synthesis (see reviews: Hopwood and Sherman, *Annu. Rev. Genet.* (1990) 24: 37-66, and Katz and Donadio, in *Annual Review of Microbiology* (1993) 47: 875-912). It has been proposed to use polyketide synthases to produce novel antibiotics (Hutchinson and Fujii, *Annual Review of Microbiology* (1995) 49:201-238).

SUMMARY OF THE INVENTION

Novel compositions and methods are provided for preparation of long chain poly-unsaturated fatty acids (PUFAs) using polyketide-like synthesis (PKS-like) genes in plants and plant cells. In contrast to the known and proposed methods for production of PUFAs by means of fatty acid synthesis genes, by the invention constructs and methods are provided for producing PUFAs by utilizing genes of a PKS-like system. The methods involve growing a host cell of interest transformed with an expression cassette functional in the host cell, the expression cassette comprising a transcriptional and translational initiation regulatory region, joined in reading frame 5' to a DNA sequence to a gene or component of a PKS-like system capable of modulating the production of PUFAs (PKS-like gene). An alteration in the PUFA profile of host cells is achieved by expression following introduction of a complete PKS-like system responsible for a PUFA biosynthesis into host cells. The invention finds use for example in the large scale production of DHA and EPA and for modification of the fatty acid profile of host cells and edible plant tissues and/or plant parts.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 provides designations for the ORFs of the EPA gene cluster of *Shewanella*. Figure 1A shows the organization of the genes; those ORFs essential for EPA production in *E. coli* are numbered. Figure 1B shows the designations given to subclones.

Figure 2 provides the *Shewanella* PKS-like domain structure, motifs and 'Blast' matches of ORF 6 (Figure 2A), ORF 7 (Figure 2B), ORF 8 (Figure 2C), ORF 9 (Figure 2D) and ORF 3 (Figure 2E). Figure 2F shows the structure of the region of the Anabeana chromosome that is related to domains present in *Shewanella* EPA ORFs.

Figure 3 shows results for pantethenylation - ORF 3 in *E. coli* strain SJ16.

Figure 4 is the sequence for the PKS-like cluster found in *Shewanella*, containing ORFs 3, 4, 5, 6, 7, 8 and 9. The start and last codons for each ORF are as follows:
ORF3 (published-inactive): 9016, 8186; ORF3 (active in EPA synthesis): 9157, 8186;
ORF 6: 13906, 22173; ORF 7: 22203, 24515; ORF 8: 24518, 30529; ORF 9: 30730, 32358.

Figure 5 shows the sequence for the PKS-like cluster in an approximately 40 kb DNA fragment of *Vibrio marinus*, containing ORFs 6, 7, 8 and 9. The start and last condons for each ORF are as follows: ORF 6: 17394, 25352; ORF 7: 25509, 28160; ORF 8: 28209, 34265; ORF 9: 34454, 36118.

Figure 6 shows the sequence for an approximately 19 kb portion of the PKS-like cluster of Figure 5 which contains the ORFs 6, 7, 8 and 9. The start and last condons for each ORF are as follows: ORF 6: 411, 8369; ORF 7: 8526, 11177; ORF 8: 11226, 17282; ORF 9: 17471, 19135.

Figure 7 shows a comparison of the PKS-like gene clusters of *Shewanella putrefaciens* and *Vibrio marinus*; Figure 7B is the *Vibrio marinus* operon sequence.

Figure 8 is an expanded view of the PKS-like gene cluster portion of *Vibrio marinus* shown in Figure 7B showing that ORFs 6, 7 and 8 are in reading frame 2, while ORF 9 is in reading frame 3.

Figure 9 demonstrates sequence homology of ORF 6 of *Shewanella putrefaciens* and *Vibrio marinus*. The *Shewanella* ORF 6 is depicted on the vertical axis, and the *Vibrio* ORF 6 is depicted on the horizontal axis. Lines indicate regions of the proteins that have a 60% identity. The repeated lines in the middle correspond to the multiple ACP domains found in ORF 6.

Figure 10 demonstrates sequence homology of ORF 7 of *Shewanella putrefaciens* and *Vibrio marinus*. The *Shewanella* ORF 7 is depicted on the vertical axis, and the *Vibrio* ORF 7 is depicted on the horizontal axis. Lines indicate regions of the proteins that have a 60% identity.

Figure 11 demonstrates sequence homology of ORF 8 of *Shewanella putrefaciens* and *Vibrio marinus*. The *Shewanella* ORF 8 is depicted on the vertical axis, and the *Vibrio* ORF 8 is depicted on the horizontal axis. Lines indicate regions of the proteins that have a 60% identity.

Figure 12 demonstrates sequence homology of ORF 9 of *Shewanella putrefaciens* and *Vibrio marinus*. The *Shewanella* ORF 9 is depicted on the vertical axis, and the *Vibrio* ORF 9 is depicted on the horizontal axis. Lines indicate regions of the proteins that have a 60% identity.

Figure 13 is a depiction of various complementation experiments, and resulting PUFA production. On the right, is shown the longest PUFA made in the *E. coli* strain

containing the *Vibrio* and *Shewanella* genes depicted on the left. The hollow boxes indicate ORFs from *Shewanella*. The solid boxes indicate ORFs from *Vibrio*.

Figure 14 is a chromatogram showing fatty acid production from complementation of pEPAD8 from *Shewanella* (deletion ORF 8) with ORF 8 from *Shewanella*, in *E. coli* Fad E-. The chromatogram presents an EPA (20:5) peak.

Figure 15 is a chromatogram showing fatty acid production from complementation of pEPAD8 from *Shewanella* (deletion ORF 8) with ORF 8 from *Vibrio marinus*, in *E. coli* Fad E-. The chromatograph presents EPA (20:5) and DHA (22:6) peaks.

Figure 16 is a table of PUFA values from the ORF 8 complementation experiment, the chromatogram of which is shown in Figure 15.

Figure 17 is a plasmid map showing the elements of pCGN7770.

Figure 18 is a plasmid map showing the elements of pCGN8535.

Figure 19 is a plasmid map showing the elements of pCGN8537.

Figure 20 is a plasmid map showing the elements of pCGN8525.

Figure 21 is a comparison of the *Shewanella* ORFs as defined by Yazawa and those disclosed in Figure 4. When a protein starting at the leucine (TTG) codon at nucleotides 9157-9155 and ending at the stop codon at nucleotides 8185-8183 is expressed under control of a heterologous promoter in an *E. coli* strain containing the entire PKS-like cluster except ORF 3, the recombinant cells do produce EPA. Thus, the published protein sequence is likely to be wrong, and the coding sequence for the protein may start at the TTG codon at nucleotides 9157-9155 or the TTG codon at nucleotides 9172-9170. This information is critical to the expression of a functional PKS-like cluster heterologous system.

Figure 22 is a plasmid map showing the elements of pCGN8560.

Figure 23 is plasmid map showing the elements of pCGN8556.

Figure 24 shows the translated DNA sequence upstream of the published ORF 3. The ATG start codon at position 9016 is the start codon for the protein described by Yazawa *et al* (1996) *supra*. The other arrows depict TTG or ATT codons that can also serve as start codons in bacteria. When ORF 3 is started from the published ATG codon at 9016, the protein is not functional in making EPA. When ORF 3 is initiated at the TTG codon at position 9157, the protein is capable of facilitating EPA synthesis.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

In accordance with the subject invention, novel DNA sequences, DNA constructs and methods are provided, which include some or all of the polyketide-like synthesis (PKS-like) pathway genes from *Shewanella*, *Vibrio* or other microorganisms, for modifying the poly-unsaturated long chain fatty acid content of host cells, particularly host plant cells. The present invention demonstrates that EPA synthesis genes in *Shewanella putrefaciens* constitute a polyketide-like synthesis pathway. Functions are ascribed to the *Shewanella* and *Vibrio* genes and methods are provided for the production of EPA and DHA in host cells. The method includes the step of transforming cells with an expression cassette comprising a DNA encoding a polypeptide capable of increasing the amount of one or more PUFA in the host cell. Desirably, integration constructs are prepared which provide for integration of the expression cassette into the genome of a host cell. Host cells are manipulated to express a sense or antisense DNA encoding a polypeptide(s) that has PKS-like gene activity. By "PKS-like gene" is intended a polypeptide which is responsible for any one or more of the functions of a PKS-like activity of interest. By "polypeptide" is meant any chain of amino acids, regardless of length or post-translational modification, for example, glycosylation or phosphorylation. Depending upon the nature of the host cell, the substrate(s) for the expressed enzyme may be produced by the host cell or may be exogenously supplied. Of particular interest is the selective control of PUFA production in plant tissues and/or plant parts such as leaves, roots, fruits and seeds. The invention can be used to synthesize EPA, DHA, and other related PUFAs in host cells.

There are many advantages to transgenic production of PUFAs. As an example, in transgenic *E. coli* as in *Shewanella*, EPA accumulates in the phospholipid fraction, specifically in the *sn*-2 position. It may be possible to produce a structured lipid in a desired host cell which differs substantially from that produced in either *Shewanella* or *E. coli*. Additionally transgenic production of PUFAs in particular host cells offers several advantages over purification from natural sources such as fish or plants. In transgenic plants, by utilizing a PKS-like system, fatty acid synthesis of PUFAs is achieved in the cytoplasm by a system which produces the PUFAs through *de novo* production of the fatty acids utilizing malonyl Co-A and acetyl Co-A as substrates. In this fashion, potential problems, such as those associated with substrate competition and diversion of normal products of fatty acid synthesis in a host to PUFA production, are avoided.

Production of fatty acids from recombinant plants provides the ability to alter the naturally occurring plant fatty acid profile by providing new synthetic pathways in the host or by suppressing undesired pathways, thereby increasing levels of desired PUFAs, or conjugated forms thereof, and decreasing levels of undesired PUFAs. Production of fatty acids in transgenic plants also offers the advantage that expression of PKS-like genes in particular tissues and/or plant parts means that greatly increased levels of desired PUFAs in those tissues and/or parts can be achieved, making recovery from those tissues more economical. Expression in a plant tissue and/or plant part presents certain efficiencies, particularly where the tissue or part is one which is easily harvested, such as seed, leaves, fruits, flowers, roots, etc. For example, the desired PUFAs can be expressed in seed; methods of isolating seed oils are well established. In addition to providing a source for purification of desired PUFAs, seed oil components can be manipulated through expression of PKS-like genes, either alone or in combination with other genes such as elongases, to provide seed oils having a particular PUFA profile in concentrated form. The concentrated seed oils then can be added to animal milks and/or synthetic or semisynthetic milks to serve as infant formulas where human nursing is impossible or undesired, or in cases of malnourishment or disease in both adults and infants.

Transgenic microbial production of fatty acids offers the advantages that many microbes are known with greatly simplified oil compositions as compared with those of higher organisms, making purification of desired components easier. Microbial production is not subject to fluctuations caused by external variables such as weather and food supply. Microbially produced oil is substantially free of contamination by environmental pollutants. Additionally, microbes can provide PUFAs in particular forms which may have specific uses. For example, *Spirulina* can provide PUFAs predominantly at the first and third positions of triglycerides; digestion by pancreatic lipases preferentially releases fatty acids from these positions. Following human or animal ingestion of triglycerides derived from *Spirulina*, these PUFAs are released by pancreatic lipases as free fatty acids and thus are directly available, for example, for infant brain development. Additionally, microbial oil production can be manipulated by controlling culture conditions, notably by providing particular substrates for microbially expressed enzymes, or by addition of compounds which suppress undesired biochemical pathways. In addition to these advantages, production of fatty acids from recombinant microbes provides the ability to alter the naturally occurring microbial fatty acid profile by

providing new synthetic pathways in the host or by suppressing undesired pathways, thereby increasing levels of desired PUFAs, or conjugated forms thereof, and decreasing levels of undesired PUFAs.

Production of fatty acids in animals also presents several advantages. Expression of desaturase genes in animals can produce greatly increased levels of desired PUFAs in animal tissues, making recovery from those tissues more economical. For example, where the desired PUFAs are expressed in the breast milk of animals, methods of isolating PUFAs from animal milk are well established. In addition to providing a source for purification of desired PUFAs, animal breast milk can be manipulated through expression of desaturase genes, either alone or in combination with other human genes, to provide animal milks with a PUFA composition substantially similar to human breast milk during the different stages of infant development. Humanized animal milks could serve as infant formulas where human nursing is impossible or undesired, or in the cases of malnourishment or disease.

DNAs encoding desired PKS-like genes can be identified in a variety of ways. In one method, a source of a desired PKS-like gene, for example genomic libraries from a *Shewanella* or *Vibrio* spp., is screened with detectable enzymatically- or chemically-synthesized probes. Sources of ORFs having PKS-like genes are those organisms which produce a desired PUFA, including DHA-producing or EPA-producing deep sea bacteria growing preferentially under high pressure or at relatively low temperature. Microorganisms such as *Shewanella* which produce EPA or DHA also can be used as a source of PKS-like genes. The probes can be made from DNA, RNA, or non-naturally occurring nucleotides, or mixtures thereof. Probes can be enzymatically synthesized from DNAs of known PKS-like genes for normal or reduced-stringency hybridization methods. For discussions of nucleic acid probe design and annealing conditions, see, for example, Sambrook *et al*, *Molecular Cloning: A Laboratory Manual* (2nd ed.), Vols. 1-3, Cold Spring Harbor Laboratory, (1989) or *Current Protocols in Molecular Biology*, F. Ausubel *et al*, ed., Greene Publishing and Wiley-Interscience, New York (1987), each of which is incorporated herein by reference. Techniques for manipulation of nucleic acids encoding PUFA enzymes such as subcloning nucleic acid sequences encoding polypeptides into expression vectors, labelling probes, DNA hybridization, and the like are described generally in Sambrook, *supra*.

Oligonucleotide probes also can be used to screen sources and can be based on sequences of known PKS-like genes, including sequences conserved among known PKS-like genes, or on peptide sequences obtained from a desired purified protein.

Oligonucleotide probes based on amino acid sequences can be degenerate to encompass the degeneracy of the genetic code, or can be biased in favor of the preferred codons of the source organism. Alternatively, a desired protein can be entirely sequenced and total synthesis of a DNA encoding that polypeptide performed.

Once the desired DNA has been isolated, it can be sequenced by known methods. It is recognized in the art that such methods are subject to errors, such that multiple sequencing of the same region is routine and is still expected to lead to measurable rates of mistakes in the resulting deduced sequence, particularly in regions having repeated domains, extensive secondary structure, or unusual base compositions, such as regions with high GC base content. When discrepancies arise, resequencing can be done and can employ special methods. Special methods can include altering sequencing conditions by using: different temperatures; different enzymes; proteins which alter the ability of oligonucleotides to form higher order structures; altered nucleotides such as ITP or methylated dGTP; different gel compositions, for example adding formamide; different primers or primers located at different distances from the problem region; or different templates such as single stranded DNAs. Sequencing of mRNA can also be employed.

For the most part, some or all of the coding sequences for the polypeptides having PKS-like gene activity are from a natural source. In some situations, however, it is desirable to modify all or a portion of the codons, for example, to enhance expression, by employing host preferred codons. Host preferred codons can be determined from the codons of highest frequency in the proteins expressed in the largest amount in a particular host species of interest. Thus, the coding sequence for a polypeptide having PKS-like gene activity can be synthesized in whole or in part. All or portions of the DNA also can be synthesized to remove any destabilizing sequences or regions of secondary structure which would be present in the transcribed mRNA. All or portions of the DNA also can be synthesized to alter the base composition to one more preferable to the desired host cell. Methods for synthesizing sequences and bringing sequences together are well established in the literature. *In vitro* mutagenesis and selection, site-directed mutagenesis, or other means can be employed to obtain mutations of naturally occurring PKS-like genes to produce a polypeptide having PKS-like gene activity *in vivo* with more desirable

physical and kinetic parameters for function in the host cell, such as a longer half-life or a higher rate of production of a desired polyunsaturated fatty acid.

Of particular interest are the *Shewanella putrefaciens* ORFs and the corresponding ORFs of *Vibrio marinus*. The *Shewanella putrefaciens* PKS-like genes can be expressed in transgenic plants to effect biosynthesis of EPA. Other DNAs which are substantially identical in sequence to the *Shewanella putrefaciens* PKS-like genes, or which encode polypeptides which are substantially similar to PKS-like genes of *Shewanella putrefaciens* can be used, such as those identified from *Vibrio marinus*. By substantially identical in sequence is intended an amino acid sequence or nucleic acid sequence exhibiting in order of increasing preference at least 60%, 80%, 90% or 95% homology to the DNA sequence of the *Shewanella putrefaciens* PKS-like genes or nucleic acid sequences encoding the amino acid sequences for such genes. For polypeptides, the length of comparison sequences generally is at least 16 amino acids, preferably at least 20 amino acids, and most preferably 35 amino acids. For nucleic acids, the length of comparison sequences generally is at least 50 nucleotides, preferably at least 60 nucleotides, and more preferably at least 75 nucleotides, and most preferably, 110 nucleotides.

Homology typically is measured using sequence analysis software, for example, the Sequence Analysis software package of the Genetics Computer Group, University of Wisconsin Biotechnology Center, 1710 University Avenue, Madison, Wisconsin 53705, MEGAlign (DNASTar, Inc., 1228 S. Park St., Madison, Wisconsin 53715), and MacVector (Oxford Molecular Group, 2105 S. Bascom Avenue, Suite 200, Campbell, California 95008). BLAST (National Center for Biotechnology Information (WCBI) www.ncbi.nlm.gov; FASTA (Pearson and Lipman, *Science* (1985) 227:1435-1446). Such software matches similar sequences by assigning degrees of homology to various substitutions, deletions, and other modifications. Conservative substitutions typically include substitutions within the following groups: glycine and alanine; valine, isoleucine and leucine; aspartic acid, glutamic acid, asparagine, and glutamine; serine and threonine; lysine and arginine; and phenylalanine and tyrosine. Substitutions may also be made on the basis of conserved hydrophobicity or hydrophilicity (Kyte and Doolittle, *J. Mol. Biol.* (1982) 157: 105-132), or on the basis of the ability to assume similar polypeptide secondary structure (Chou and Fasman, *Adv. Enzymol.* (1978) 47: 45-148, 1978). A

related protein to the probing sequence is identified when $p \geq 0.01$, preferably $p \geq 10^{-7}$ or 10^{-8} .

Encompassed by the present invention are related PKS-like genes from the same or other organisms. Such related PKS-like genes include variants of the disclosed PKS-like ORFs that occur naturally within the same or different species of *Shewanella*, as well as homologues of the disclosed PKS-like genes from other species and evolutionarily related proteins having analogous function and activity. Also included are PKS-like genes which, although not substantially identical to the *Shewanella putrefaciens* PKS-like genes, operate in a similar fashion to produce PUFAs as part of a PKS-like system.

Related PKS-like genes can be identified by their ability to function substantially the same as the disclosed PKS-like genes; that is, they can be substituted for corresponding ORFs of *Shewanella* or *Vibrio* and still effectively produce EPA or DHA. Related PKS-like genes also can be identified by screening sequence databases for sequences homologous to the disclosed PKS-like genes, by hybridization of a probe based on the disclosed PKS-like genes to a library constructed from the source organism, or by RT-PCR using mRNA from the source organism and primers based on the disclosed PKS-like gene. Thus, the phrase "PKS-like genes" refers not only to the nucleotide sequences disclosed herein, but also to other nucleic acids that are allelic or species variants of these nucleotide sequences. It is also understood that these terms include nonnatural mutations introduced by deliberate mutation using recombinant technology such as single site mutation or by excising short sections of DNA open reading frames coding for PUFA enzymes or by substituting new codons or adding new codons. Such minor alterations substantially maintain the immunoidentity of the original expression product and/or its biological activity. The biological properties of the altered PUFA enzymes can be determined by expressing the enzymes in an appropriate cell line and by determining the ability of the enzymes to synthesize PUFAs. Particular enzyme modifications considered minor would include substitution of amino acids of similar chemical properties, e.g., glutamic acid for aspartic acid or glutamine for asparagine.

When utilizing a PUFA PKS-like system from another organism, the regions of a PKS-like gene polypeptide important for PKS-like gene activity can be determined through routine mutagenesis, expression of the resulting mutant polypeptides and determination of their activities. The coding region for the mutants can include deletions, insertions and point mutations, or combinations thereof. A typical functional analysis

begins with deletion mutagenesis to determine the N- and C-terminal limits of the protein necessary for function, and then internal deletions, insertions or point mutants are made in the open ready frame to further determine regions necessary for function. Other techniques such as cassette mutagenesis or total synthesis also can be used. Deletion mutagenesis is accomplished, for example, by using exonucleases to sequentially remove the 5' or 3' coding regions. Kits are available for such techniques. After deletion, the coding region is completed by ligating oligonucleotides containing start or stop codons to the deleted coding region after 5' or 3' deletion, respectively. Alternatively, oligonucleotides encoding start or stop codons are inserted into the coding region by a variety of methods including site-directed mutagenesis, mutagenic PCR or by ligation onto DNA digested at existing restriction sites. Internal deletions can similarly be made through a variety of methods including the use of existing restriction sites in the DNA, by use of mutagenic primers via site directed mutagenesis or mutagenic PCR. Insertions are made through methods such as linker-scanning mutagenesis, site-directed mutagenesis or mutagenic PCR. Point mutations are made through techniques such as site-directed mutagenesis or mutagenic PCR.

Chemical mutagenesis also can be used for identifying regions of a PKS-like gene polypeptide important for activity. A mutated construct is expressed, and the ability of the resulting altered protein to function as a PKS-like gene is assayed. Such structure-function analysis can determine which regions may be deleted, which regions tolerate insertions, and which point mutations allow the mutant protein to function in substantially the same way as the native PKS-like gene. All such mutant proteins and nucleotide sequences encoding them are within the scope of the present invention. EPA is produced in *Shewanella* as the product of a PKS-like system, such that the EPA genes encode components of this system. In *Vibrio*, DHA is produced by a similar system. The enzymes which synthesize these fatty acids are encoded by a cluster of genes which are distinct from the fatty acid synthesis genes encoding the enzymes involved in synthesis of the C16 and C18 fatty acids typically found in bacteria and in plants. As the *Shewanella* EPA genes represent a PKS-like gene cluster, EPA production is, at least to some extent, independent of the typical bacterial type II FAS system. Thus, production of EPA in the cytoplasm of plant cells can be achieved by expression of the PKS-like pathway genes in plant cells under the control of appropriate plant regulatory signals.

EPA production in *E. coli* transformed with the *Shewanella* EPA genes proceeds during anaerobic growth, indicating that O₂-dependent desaturase reactions are not involved. Analyses of the proteins encoded by the ORFs essential for EPA production reveals the presence of domain structures characteristic of PKS-like systems. Fig. 2A shows a summary of the domains, motifs, and also key homologies detected by "BLAST" data bank searches. Because EPA is different from many of the other substances produced by PKS-like pathways, i.e., it contains 5, *cis* double bonds, spaced at 3 carbon intervals along the molecule, a PKS-like system for synthesis of EPA is not expected.

Further, BLAST searches using the domains present in the *Shewanella* EPA ORFs reveal that several are related to proteins encoded by a PKS-like gene cluster found in Anabeana. The structure of that region of the Anabeana chromosome is shown in Fig. 2F. The Anabeana PKS-like genes have been linked to the synthesis of a long-chain (C₂₆), hydroxy-fatty acid found in a glycolipid layer of heterocysts. The EPA protein domains with homology to the Anabeana proteins are indicated in Fig. 2F.

ORF 6 of *Shewanella* contains a KAS domain which includes an active site motif (DXAC*) as well as a "GFGG" motif which is present at the end of many Type II KAS proteins (see Fig. 2A). Extended motifs are present but not shown here. Next is a malonyl-CoA:ACP acyl transferase (AT) domain. Sequences near the active site motif (GHS**XG*) suggest it transfers malonate rather than methylmalonate, i.e., it resembles the acetate-like ATs. Following a linker region, there is a cluster of 6 repeating domains, each ~100 amino acids in length, which are homologous to PKS-like ACP sequences. Each contains a pantetheine binding site motif (LGXDS*(L/I)). The presence of 6 such ACP domains has not been observed previously in fatty acid synthases (FAS) or PKS-like systems. Near the end of the protein is a region which shows homology to β -keto-ACP reductases (KR). It contains a pyridine nucleotide binding site motif "GXGXX(G/A/P)".

The *Shewanella* ORF 8 begins with a KAS domain, including active site and ending motifs (Fig. 2C). The best match in the data banks is with the Anabeana HglD. There is also a domain which has sequence homology to the N-terminal one half of the Anabeana HglC. This region also shows weak homology to KAS proteins although it lacks the active site and ending motifs. It has the characteristics of the so-called chain length factors (CLF) of Type II PKS-like systems. ORF 8 appears to direct the production of EPA versus DHA by the PKS-like system. ORF 8 also has two domains with homology to β -hydroxyacyl-ACP dehydrases (DH). The best match for both domains is

with *E. coli* FabA, a bi-functional enzyme which carries out both the dehydrase reaction and an isomerization (*trans* to *cis*) of the resulting double bond. The first DH domain contains both the active site histidine (H) and an adjacent cysteine (C) implicated in FabA catalysis. The second DH domain has the active site H but lacks the adjacent C (Fig. 2C).

5 Blast searches with the second DH domain also show matches to FabZ, a second *E. coli* DH, which does not possess isomerase activity.

The N-terminal half of ORF 7 (Fig. 2B) has no significant matches in the data banks. The best match of the C-terminal half is with a C-terminal portion of the Anabeana HglC. This domain contains an acyl-transferase (AT) motif (GX SXG).

10 Comparison of the extended active site sequences, based on the crystal structure of the *E. coli* malonyl-CoA:ACP AT, reveals that ORF 7 lacks two residues essential for exclusion of water from the active site (*E. coli* nomenclature; Q11 and R117). These data suggest that ORF 7 may function as a thioesterase.

ORF 9 (Fig. 2D) is homologous to an ORF of unknown function in the Anabeana Hgl cluster. It also exhibits a very weak homology to NIFA, a regulatory protein in nitrogen fixing bacteria. A regulatory role for the ORF 9 protein has not been excluded. ORF 3 (Fig. 2E) is homologous to the Anabeana HetI as well as EntD from *E. coli* and Sfp of *Bacillus*. Recently, a new enzyme family of phosphopantetheinyl transferases has been identified that includes HetI, EntD and Sfp (Lamblot RH, *et al.* (1996) A new enzyme superfamily - the phosphopantetheinyl transferases. *Chemistry & Biology*, Vol 3, #11, 923-936). The data of Fig. 3 demonstrates that the presence of ORF 3 is required for addition of β -alanine (i.e. pantetheine) to the ORF 6 protein. Thus, ORF 3 encodes the phosphopantetheinyl transferase specific for the ORF 6 ACP domains. (See, Haydock SF *et al.* (1995) Divergent sequence motifs correlated with the substrate specificity of (methyl)malonyl-CoA:acyl carrier protein transacylase domains in modular polyketide synthases, *FEBS Lett.*, 374, 246-248). Malonate is the source of the carbons utilized in the extension reactions of EPA synthesis. Additionally, malonyl-CoA rather than malonyl-ACP is the AT substrate, i.e., the AT region of ORF 6 uses malonyl Co-A.

Once the DNA sequences encoding the PKS-like genes of an organism responsible for PUFA production have been obtained, they are placed in a vector capable of replication in a host cell, or propagated *in vitro* by means of techniques such as PCR or long PCR. Replicating vectors can include plasmids, phage, viruses, cosmids and the like. Desirable vectors include those useful for mutagenesis of the gene of interest or for

expression of the gene of interest in host cells. A PUFA synthesis enzyme or a homologous protein can be expressed in a variety of recombinantly engineered cells. Numerous expression systems are available for expression of DNA encoding a PUFA enzyme. The expression of natural or synthetic nucleic acids encoding PUFA enzyme is typically achieved by operably linking the DNA to a promoter (which is either constitutive or inducible) within an expression vector. By expression vector is meant a DNA molecule, linear or circular, that comprises a segment encoding a PUFA enzyme, operably linked to additional segments that provide for its transcription. Such additional segments include promoter and terminator sequences. An expression vector also may include one or more origins of replication, one or more selectable markers, an enhancer, a polyadenylation signal, etc. Expression vectors generally are derived from plasmid or viral DNA, and can contain elements of both. The term "operably linked" indicates that the segments are arranged so that they function in concert for their intended purposes, for example, transcription initiates in the promoter and proceeds through the coding segment to the terminator. See Sambrook *et al*, *supra*.

The technique of long PCR has made *in vitro* propagation of large constructs possible, so that modifications to the gene of interest, such as mutagenesis or addition of expression signals, and propagation of the resulting constructs can occur entirely *in vitro* without the use of a replicating vector or a host cell. *In vitro* expression can be accomplished, for example, by placing the coding region for the desaturase polypeptide in an expression vector designed for *in vitro* use and adding rabbit reticulocyte lysate and cofactors; labeled amino acids can be incorporated if desired. Such *in vitro* expression vectors may provide some or all of the expression signals necessary in the system used. These methods are well known in the art and the components of the system are commercially available. The reaction mixture can then be assayed directly for PKS-like enzymes for example by determining their activity, or the synthesized enzyme can be purified and then assayed.

Expression in a host cell can be accomplished in a transient or stable fashion. Transient expression can occur from introduced constructs which contain expression signals functional in the host cell, but which constructs do not replicate and rarely integrate in the host cell, or where the host cell is not proliferating. Transient expression also can be accomplished by inducing the activity of a regulatable promoter operably linked to the gene of interest, although such inducible systems frequently exhibit a low

basal level of expression. Stable expression can be achieved by introduction of a nucleic acid construct that can integrate into the host genome or that autonomously replicates in the host cell. Stable expression of the gene of interest can be selected for through the use of a selectable marker located on or transfected with the expression construct, followed by selection for cells expressing the marker. When stable expression results from integration, integration of constructs can occur randomly within the host genome or can be targeted through the use of constructs containing regions of homology with the host genome sufficient to target recombination with the host locus. Where constructs are targeted to an endogenous locus, all or some of the transcriptional and translational regulatory regions can be provided by the endogenous locus. To achieve expression in a host cell, the transformed DNA is operably associated with transcriptional and translational initiation and termination regulatory regions that are functional in the host cell.

Transcriptional and translational initiation and termination regions are derived from a variety of nonexclusive sources, including the DNA to be expressed, genes known or suspected to be capable of expression in the desired system, expression vectors, chemical synthesis. The termination region can be derived from the 3' region of the gene from which the initiation region was obtained or from a different gene. A large number of termination regions are known to and have been found to be satisfactory in a variety of hosts from the same and different genera and species. The termination region usually is selected more as a matter of convenience rather than because of any particular property. When expressing more than one PKS-like ORF in the same cell, appropriate regulatory regions and expression methods should be used. Introduced genes can be propagated in the host cell through use of replicating vectors or by integration into the host genome. Where two or more genes are expressed from separate replicating vectors, it is desirable that each vector has a different means of replication. Each introduced construct, whether integrated or not, should have a different means of selection and should lack homology to the other constructs to maintain stable expression and prevent reassortment of elements among constructs. Judicious choices of regulatory regions, selection means and method of propagation of the introduced construct can be experimentally determined so that all introduced genes are expressed at the necessary levels to provide for synthesis of the desired products.

A variety of procaryotic expression systems can be used to express PUFA enzyme. Expression vectors can be constructed which contain a promoter to direct transcription, a ribosome binding site, and a transcriptional terminator. Examples of regulatory regions suitable for this purpose in *E. coli* are the promoter and operator region of the *E. coli* tryptophan biosynthetic pathway as described by Yanofsky (1984) *J. Bacteriol.*, 158:1018-1024 and the leftward promoter of phage lambda (P λ) as described by Herskowitz and Hagen, (1980) *Ann. Rev. Genet.*, 14:399-445. The inclusion of selection markers in DNA vectors transformed in *E. coli* is also useful. Examples of such markers include genes specifying resistance to ampicillin, tetracycline, or chloramphenicol.

Vectors used for expressing foreign genes in bacterial hosts generally will contain a selectable marker, such as a gene for antibiotic resistance, and a promoter which functions in the host cell. Plasmids useful for transforming bacteria include pBR322 (Bolivar, *et al*, (1977) *Gene* 2:95-113), the pUC plasmids (Messing, (1983) *Meth. Enzymol.* 101:20-77, Vieira and Messing, (1982) *Gene* 19:259-268), pCQV2 (Queen, *ibid.*), and derivatives thereof. Plasmids may contain both viral and bacterial elements. Methods for the recovery of the proteins in biologically active form are discussed in U.S. Patent Nos. 4,966,963 and 4,999,422, which are incorporated herein by reference. See Sambrook, *et al* for a description of other prokaryotic expression systems.

For expression in eukaryotes, host cells for use in practicing the present invention include mammalian, avian, plant, insect, and fungal cells. As an example, for plants, the choice of a promoter will depend in part upon whether constitutive or inducible expression is desired and whether it is desirable to produce the PUFAs at a particular stage of plant development and/or in a particular tissue. Considerations for choosing a specific tissue and/or developmental stage for expression of the ORFs may depend on competing substrates or the ability of the host cell to tolerate expression of a particular PUFA. Expression can be targeted to a particular location within a host plant such as seed, leaves, fruits, flowers, and roots, by using specific regulatory sequences, such as those described in USPN 5,463,174, USPN 4,943,674, USPN 5,106,739, USPN 5,175,095, USPN 5,420,034, USPN 5,188,958, and USPN 5,589,379. Where the host cell is a yeast, transcription and translational regions functional in yeast cells are provided, particularly from the host species. The transcriptional initiation regulatory regions can be obtained, for example from genes in the glycolytic pathway, such as alcohol dehydrogenase, glyceraldehyde-3-phosphate dehydrogenase (GPD),

phosphoglucosomerase; phosphoglycerate kinase, etc. or regulatable genes such as acid phosphatase, lactase, metallothionein, glucoamylase, etc. Any one of a number of regulatory sequences can be used in a particular situation, depending upon whether constitutive or induced transcription is desired, the particular efficiency of the promoter in conjunction with the open-reading frame of interest, the ability to join a strong promoter with a control region from a different promoter which allows for inducible transcription, ease of construction, and the like. Of particular interest are promoters which are activated in the presence of galactose. Galactose-inducible promoters (GAL1, GAL7, and GAL10) have been extensively utilized for high level and regulated expression of protein in yeast (Lue *et al*, (1987) *Mol. Cell. Biol.* 7:3446; Johnston, (1987) *Microbiol. Rev.* 51:458).

Transcription from the GAL promoters is activated by the GAL4 protein, which binds to the promoter region and activates transcription when galactose is present. In the absence of galactose, the antagonist GAL80 binds to GAL4 and prevents GAL4 from activating transcription. Addition of galactose prevents GAL80 from inhibiting activation by GAL4. Preferably, the termination region is derived from a yeast gene, particularly *Saccharomyces*, *Schizosaccharomyces*, *Candida* or *Kluyveromyces*. The 3' regions of two mammalian genes, γ interferon and $\alpha 2$ interferon, are also known to function in yeast.

Nucleotide sequences surrounding the translational initiation codon ATG have been found to affect expression in yeast cells. If the desired polypeptide is poorly expressed in yeast, the nucleotide sequences of exogenous genes can be modified to include an efficient yeast translation initiation sequence to obtain optimal gene expression. For expression in *Saccharomyces*, this can be done by site-directed mutagenesis of an inefficiently expressed gene by fusing it in-frame to an endogenous *Saccharomyces* gene, preferably a highly expressed gene, such as the lactase gene.

As an alternative to expressing the PKS-like genes in the plant cell cytoplasm, is to target the enzymes to the chloroplast. One method to target proteins to the chloroplast entails use of leader peptides attached to the N-termini of the proteins. Commonly used leader peptides are derived from the small subunit of plant ribulose bis phosphate carboxylase. Leader sequences from other chloroplast proteins may also be used.

Another method for targeting proteins to the chloroplast is to transform the chloroplast genome (Stable transformation of chloroplasts of *Chlamydomonas reinhardtii* (1 green alga) using bombardment of recipient cells with high-velocity tungsten microprojectiles coated with foreign DNA has been described. See, for example, Blowers *et al Plant Cell*

(1989) 1:123-132 and Debuchy *et al EMBO J* (1989) 8:2803-2809. The transformation technique, using tungsten microprojectiles, is described by Kline *et al, Nature* (London) (1987) 327:70-73). The most common method of transforming chloroplasts involves using biolistic techniques, but other techniques developed for the purpose may also be used. (Methods for targeting foreign gene products into chloroplasts (Shrier *et al EMBO J.* (1985) 4:25-32) or mitochondria (Boutry *et al, supra*) have been described. See also Tomai *et al Gen. Biol. Chem.* (1988) 263:15104-15109 and US Patent No. 4,940,835 for the use of transit peptides for translocating nuclear gene products into the chloroplast. Methods for directing the transport of proteins to the chloroplast are reviewed in Kenauf
10 *TIBTECH* (1987) 5:40-47.

For producing PUFAs in avian species and cells, gene transfer can be performed by introducing a nucleic acid sequence encoding a PUFA enzyme into the cells following procedures known in the art. If a transgenic animal is desired, pluripotent stem cells of embryos can be provided with a vector carrying a PUFA enzyme encoding transgene and
15 developed into adult animal (USPN 5,162,215; Ono *et al.* (1996) *Comparative Biochemistry and Physiology A* 113(3):287-292; WO 9612793; WO 9606160). In most cases, the transgene is modified to express high levels of the PKS-like enzymes in order to increase production of PUFAs. The transgenes can be modified, for example, by providing transcriptional and/or translational regulatory regions that function in avian
20 cells, such as promoters which direct expression in particular tissues and egg parts such as yolk. The gene regulatory regions can be obtained from a variety of sources, including chicken anemia or avian leukosis viruses or avian genes such as a chicken ovalbumin gene.

Production of PUFAs in insect cells can be conducted using baculovirus
25 expression vectors harboring PKS-like transgenes. Baculovirus expression vectors are available from several commercial sources such as Clontech. Methods for producing hybrid and transgenic strains of algae, such as marine algae, which contain and express a desaturase transgene also are provided. For example, transgenic marine algae can be prepared as described in USPN 5,426,040. As with the other expression systems
30 described above, the timing, extent of expression and activity of the desaturase transgene can be regulated by fitting the polypeptide coding sequence with the appropriate transcriptional and translational regulatory regions selected for a particular use. Of particular interest are promoter regions which can be induced under preselected growth

conditions. For example, introduction of temperature sensitive and/or metabolite responsive mutations into the desaturase transgene coding sequences, its regulatory regions, and/or the genome of cells into which the transgene is introduced can be used for this purpose.

5 The transformed host cell is grown under appropriate conditions adapted for a desired end result. For host cells grown in culture, the conditions are typically optimized to produce the greatest or most economical yield of PUFAs, which relates to the selected desaturase activity. Media conditions which may be optimized include: carbon source,
10 nitrogen source, addition of substrate, final concentration of added substrate, form of substrate added, aerobic or anaerobic growth, growth temperature, inducing agent, induction temperature, growth phase at induction, growth phase at harvest, pH, density, and maintenance of selection. Microorganisms such as yeast, for example, are preferably grown using selected media of interest, which include yeast peptone broth (YPD) and minimal media (contains amino acids, yeast nitrogen base, and ammonium sulfate, and
15 lacks a component for selection, for example uracil). Desirably, substrates to be added are first dissolved in ethanol. Where necessary, expression of the polypeptide of interest may be induced, for example by including or adding galactose to induce expression from a GAL promoter.

20 When increased expression of the PKS-like gene polypeptide in a host cell which expresses PUFA from a PKS-like system is desired, several methods can be employed. Additional genes encoding the PKS-like gene polypeptide can be introduced into the host organism. Expression from the native PKS-like gene locus also can be increased through homologous recombination, for example by inserting a stronger promoter into the host genome to cause increased expression, by removing destabilizing sequences from either
25 the mRNA or the encoded protein by deleting that information from the host genome, or by adding stabilizing sequences to the mRNA (*see* USPN 4,910,141 and USPN 5,500,365). Thus, the subject host will have at least have one copy of the expression construct and may have two or more, depending upon whether the gene is integrated into the genome, amplified, or is present on an extrachromosomal element having multiple
30 copy numbers. Where the subject host is a yeast, four principal types of yeast plasmid vectors can be used: Yeast Integrating plasmids (YIps), Yeast Replicating plasmids (YRps), Yeast Centromere plasmids (YCps), and Yeast Episomal plasmids (YEps). YIps lack a yeast replication origin and must be propagated as integrated elements in the yeast

genome. YRps have a chromosomally derived autonomously replicating sequence and are propagated as medium copy number (20 to 40), autonomously replicating, unstably segregating plasmids. YCps have both a replication origin and a centromere sequence and propagate as low copy number (10-20), autonomously replicating, stably segregating plasmids. YEp have an origin of replication from the yeast 2 μ m plasmid and are propagated as high copy number, autonomously replicating, irregularly segregating plasmids. The presence of the plasmids in yeast can be ensured by maintaining selection for a marker on the plasmid. Of particular interest are the yeast vectors pYES2 (a YEp plasmid available from Invitrogen, confers uracil prototrophy and a GAL1 galactose-inducible promoter for expression), and pYX424 (a YEp plasmid having a constitutive TP1 promoter and conferring leucine prototrophy; (Alber and Kawasaki (1982). *J. Mol. & Appl. Genetics* 1: 419).

The choice of a host cell is influenced in part by the desired PUFA profile of the transgenic cell, and the native profile of the host cell. Even where the host cell expresses PKS-like gene activity for one PUFA, expression of PKS-like genes of another PKS-like system can provide for production of a novel PUFA not produced by the host cell. In particular instances where expression of PKS-like gene activity is coupled with expression of an ORF 8 PKS-like gene of an organism which produces a different PUFA, it can be desirable that the host cell naturally have, or be mutated to have, low PKS-like gene activity for ORF 8. As an example, for production of EPA, the DNA sequence used encodes the polypeptide having PKS-like gene activity of an organism which produces EPA, while for production of DHA, the DNA sequences used are those from an organism which produces DHA. For use in a host cell which already expresses PKS-like gene activity it can be necessary to utilize an expression cassette which provides for overexpression of the desired PKS-like genes alone or with a construct to downregulate the activity of an existing ORF of the existing PKS-like system, such as by antisense or co-suppression. Similarly, a combination of ORFs derived from separate organisms which produce the same or different PUFAs using PKS-like systems may be used. For instance, the ORF 8 of *Vibrio* directs the expression of DHA in a host cell, even when ORFs 3, 6, 7 and 9 are from *Shewanella*, which produce EPA when coupled to ORF 8 of *Shewanella*. Therefore, for production of eicosapentanoic acid (EPA), the expression cassettes used generally include one or more cassettes which include ORFs 3, 6, 7, 8 and 9 from a PUFA-producing organism such as the marine bacterium *Shewanella*.

putrefaciens (for EPA production) or *Vibrio marinus* (for DHA production). ORF 8 can be used for induction of DHA production, and ORF 8 of *Vibrio* can be used in conjunction with ORFs 3, 6, 7 and 9 of *Shewanella* to produce DHA. The organization and numbering scheme of the ORFs identified in the *Shewanella* gene cluster are shown in Fig 1A. Maps of several subclones referred to in this study are shown in Fig 1B. For expression of a PKS-like gene polypeptide, transcriptional and translational initiation and termination regions functional in the host cell are operably linked to the DNA encoding the PKS-like gene polypeptide.

Constructs comprising the PKS-like ORFs of interest can be introduced into a host cell by any of a variety of standard techniques, depending in part upon the type of host cell. These techniques include transfection, infection, bolistic impact, electroporation, microinjection, scraping, or any other method which introduces the gene of interest into the host cell (*see* USPN 4,743,548, USPN 4,795,855, USPN 5,068,193, USPN 5,188,958, USPN 5,463,174, USPN 5,565,346 and USPN 5,565,347). Methods of transformation which are used include lithium acetate transformation (*Methods in Enzymology*, (1991) 194:186-187). For convenience, a host cell which has been manipulated by any method to take up a DNA sequence or construct will be referred to as "transformed" or "recombinant" herein. The subject host will have at least have one copy of the expression construct and may have two or more, depending upon whether the gene is integrated into the genome, amplified, or is present on an extrachromosomal element having multiple copy numbers.

For production of PUFAs, depending upon the host cell, the several polypeptides produced by pEPA, ORFs 3, 6, 7, 8 and 9, are introduced as individual expression constructs or can be combined into two or more cassettes which are introduced individually or co-transformed into a host cell. A standard transformation protocol is used. For plants, where less than all PKS-like genes required for PUFA synthesis have been inserted into a single plant, plants containing a complementing gene or genes can be crossed to obtain plants containing a full complement of PKS-like genes to synthesize a desired PUFA.

The PKS-like-mediated production of PUFAs can be performed in either prokaryotic or eukaryotic host cells. The cells can be cultured or formed as part or all of a host organism including an animal. Viruses and bacteriophage also can be used with appropriate cells in the production of PUFAs, particularly for gene transfer, cellular

targeting and selection. Any type of plant cell can be used for host cells, including dicotyledonous plants, monocotyledonous plants, and cereals. Of particular interest are crop plants such as *Brassica*, *Arabidopsis*, soybean, corn, and the like. Prokaryotic cells of interest include *Eschericia*, *Baccillus*, *Lactobaccillus*, *cyanobacteria* and the like.

5 Eukaryotic cells include plant cells, mammalian cells such as those of lactating animals, avian cells such as of chickens, and other cells amenable to genetic manipulation including insect, fungal, and algae cells. Examples of host animals include mice, rats, rabbits, chickens, quail, turkeys, cattle, sheep, pigs, goats, yaks, etc., which are amenable to genetic manipulation and cloning for rapid expansion of a transgene expressing
10 population. For animals, PKS-like transgenes can be adapted for expression in target organelles, tissues and body fluids through modification of the gene regulatory regions. Of particular interest is the production of PUFAs in the breast milk of the host animal.

Examples of host microorganisms include *Saccharomyces cerevisiae*,
Saccharomyces carlsbergensis, or other yeast such as *Candida*, *Kluyveromyces* or other
15 fungi, for example, filamentous fungi such as *Aspergillus*, *Neurospora*, *Penicillium*, etc. Desirable characteristics of a host microorganism are, for example, that it is genetically well characterized, can be used for high level expression of the product using ultra-high density fermentation, and is on the GRAS (generally recognized as safe) list since the proposed end product is intended for ingestion by humans. Of particular interest is use of
20 a yeast, more particularly baker's yeast (*S. cerevisiae*), as a cell host in the subject invention. Strains of particular interest are SC334 (Mat α pep4-3 prbl-1122 ura3-52 leu2-3, 112 reg1-501 gal1; (Hovland *et al* (1989) Gene 83:57-64); BJ1995 (Yeast Genetic Stock Centre, 1021 Donner Laboratory, Berkeley, CA 94720), INVSC1 (Mat α hiw3 Δ 1 leu2 trp1-289 ura3-52 (Invitrogen, 1600 Faraday Ave., Carlsbad, CA 92008) and INVSC2
25 (Mat α his3 Δ 200 ura3-167; (Invitrogen). Bacterial cells also may be used as hosts. This includes *E. coli*, which can be useful in fermentation processes. Alternatively, a host such as a *Lactobacillus* species can be used as a host for introducing the products of the PKS-like pathway into a product such as yogurt.

The transformed host cell can be identified by selection for a marker contained on
30 the introduced construct. Alternatively, a separate marker construct can be introduced with the desired construct, as many transformation techniques introduce multiple DNA molecules into host cells. Typically, transformed hosts are selected for their ability to grow on selective media. Selective media can incorporate an antibiotic or lack a factor

necessary for growth of the untransformed host, such as a nutrient or growth factor. An introduced marker gene therefor may confer antibiotic resistance, or encode an essential growth factor or enzyme, and permit growth on selective media when expressed in the transformed host cell. Desirably, resistance to kanamycin and the amino glycoside G418 are of particular interest (*see* USPN 5,034,322). For yeast transformants, any marker that functions in yeast can be used, such as the ability to grow on media lacking uracil, lencine, lysine or tryptophan.

Selection of a transformed host also can occur when the expressed marker protein can be detected, either directly or indirectly. The marker protein can be expressed alone or as a fusion to another protein. The marker protein can be one which is detected by its enzymatic activity; for example β -galactosidase can convert the substrate X-gal to a colored product, and luciferase can convert luciferin to a light-emitting product. The marker protein can be one which is detected by its light-producing or modifying characteristics; for example, the green fluorescent protein of *Aequorea victoria* fluoresces when illuminated with blue light. Antibodies can be used to detect the marker protein or a molecular tag on, for example, a protein of interest. Cells expressing the marker protein or tag can be selected, for example, visually, or by techniques such as FACS or panning using antibodies.

The PUFAs produced using the subject methods and compositions are found in the host plant tissue and/or plant part as free fatty acids and/or in conjugated forms such as acylglycerols, phospholipids, sulfolipids or glycolipids, and can be extracted from the host cell through a variety of means well-known in the art. Such means include extraction with organic solvents, sonication, supercritical fluid extraction using for example carbon dioxide, and physical means such as presses, or combinations thereof. Of particular interest is extraction with methanol and chloroform. Where appropriate, the aqueous layer can be acidified to protonate negatively charged moieties and thereby increase partitioning of desired products into the organic layer. After extraction, the organic solvents can be removed by evaporation under a stream of nitrogen. When isolated in conjugated forms, the products are enzymatically or chemically cleaved to release the free fatty acid or a less complex conjugate of interest, and are then subjected to further manipulations to produce a desired end product. Desirably, conjugated forms of fatty acids are cleaved with potassium hydroxide.

If further purification is necessary, standard methods can be employed. Such methods include extraction, treatment with urea, fractional crystallization, HPLC, fractional distillation, silica gel chromatography, high speed centrifugation or distillation, or combinations of these techniques. Protection of reactive groups, such as the acid or alkenyl groups, can be done at any step through known techniques, for example alkylation or iodination. Methods used include methylation of the fatty acids to produce methyl esters. Similarly, protecting groups can be removed at any step. Desirably, purification of fractions containing DHA and EPA is accomplished by treatment with urea and/or fractional distillation.

The uses of the subject invention are several. Probes based on the DNAs of the present invention find use in methods for isolating related molecules or in methods to detect organisms expressing PKS-like genes. When used as probes, the DNAs or oligonucleotides need to be detectable. This is usually accomplished by attaching a label either at an internal site, for example via incorporation of a modified residue, or at the 5' or 3' terminus. Such labels can be directly detectable, can bind to a secondary molecule that is detectably labeled, or can bind to an unlabelled secondary molecule and a detectably labeled tertiary molecule; this process can be extended as long as is practicable to achieve a satisfactorily detectable signal without unacceptable levels of background signal. Secondary, tertiary, or bridging systems can include use of antibodies directed against any other molecule, including labels or other antibodies, or can involve any molecules which bind to each other, for example a biotin-streptavidin/avidin system. Detectable labels typically include radioactive isotopes, molecules which chemically or enzymatically produce or alter light, enzymes which produce detectable reaction products, magnetic molecules, fluorescent molecules or molecules whose fluorescence or light-emitting characteristics change upon binding. Examples of labelling methods can be found in USPN 5,011,770. Alternatively, the binding of target molecules can be directly detected by measuring the change in heat of solution on binding of a probe to a target via isothermal titration calorimetry, or by coating the probe or target on a surface and detecting the change in scattering of light from the surface produced by binding of a target or a probe, respectively, is done with the BIAcore system.

PUFAs produced by recombinant means find applications in a wide variety of areas. Supplementation of humans or animals with PUFAs in various forms can result in increased levels not only of the added PUFAs, but of their metabolic progeny as well.

Complex regulatory mechanisms can make it desirable to combine various PUFAs, or to add different conjugates of PUFAs, in order to prevent, control or overcome such mechanisms to achieve the desired levels of specific PUFAs in an individual. In the present case, expression of PKS-like gene genes, or antisense PKS-like gene transcripts, can alter the levels of specific PUFAs, or derivatives thereof, found in plant parts and/or plant tissues. The PKS-like gene polypeptide coding region is expressed either by itself or with other genes, in order to produce tissues and/or plant parts containing higher proportions of desired PUFAs or containing a PUFA composition which more closely resembles that of human breast milk (Prieto *et al.*, PCT publication WO 95/24494) than does the unmodified tissues and/or plant parts.

PUFAs, or derivatives thereof, made by the disclosed method can be used as dietary supplements for patients undergoing intravenous feeding or for preventing or treating malnutrition. For dietary supplementation, the purified PUFAs, or derivatives thereof, can be incorporated into cooking oils, fats or margarines formulated so that in normal use the recipient receives a desired amount of PUFA. The PUFAs also can be incorporated into infant formulas, nutritional supplements or other food products, and find use as anti-inflammatory or cholesterol lowering agents.

Particular fatty acids such as EPA can be used to alter the composition of infant formulas to better replicate the PUFA composition of human breast milk. The predominant triglyceride in human milk is reported to be 1,3-di-oleoyl-2-palmitoyl, with 2-palmitoyl glycerides reported as better absorbed than 2-oleoyl or 2-lineoyl glycerides (*see* USPN 4,876,107). Typically, human breast milk has a fatty acid profile comprising from about 0.15 % to about 0.36 % as DHA, from about 0.03 % to about 0.13 % as EPA, from about 0.30 % to about 0.88 % as ARA, from about 0.22 % to about 0.67 % as DGLA, and from about 0.27 % to about 1.04 % as GLA. A preferred ratio of GLA:DGLA:ARA in infant formulas is from about 1:1:4 to about 1:1:1, respectively. Amounts of oils providing these ratios of PUFA can be determined without undue experimentation by one of skill in the art. PUFAs, or host cells containing them, also can be used as animal food supplements to alter an animal's tissue or milk fatty acid composition to one more desirable for human or animal consumption.

For pharmaceutical use (human or veterinary), the compositions generally are administered orally but can be administered by any route by which they may be successfully absorbed, e.g., parenterally (i.e. subcutaneously, intramuscularly or

intravenously), rectally or vaginally or topically, for example, as a skin ointment or lotion. Where available, gelatin capsules are the preferred form of oral administration. Dietary supplementation as set forth above also can provide an oral route of administration. The unsaturated acids of the present invention can be administered in conjugated forms, or as salts, esters, amides or prodrugs of the fatty acids. Any pharmaceutically acceptable salt is encompassed by the present invention; especially preferred are the sodium, potassium or lithium salts. Also encompassed are the N-alkylpolyhydroxamine salts, such as N-methyl glucamine, described in PCT publication WO 96/33155. Preferred esters are the ethyl esters.

The PUFAs of the present invention can be administered alone or in combination with a pharmaceutically acceptable carrier or excipient. As solid salts, the PUFAs can also be administered in tablet form. For intravenous administration, the PUFAs or derivatives thereof can be incorporated into commercial formulations such as Intralipids. Where desired, the individual components of formulations can be individually provided in kit form, for single or multiple use. A typical dosage of a particular fatty acid is from 0.1 mg to 20 g, or even 100 g daily, and is preferably from 10 mg to 1, 2, 5 or 10 g daily as required, or molar equivalent amounts of derivative forms thereof. Parenteral nutrition compositions comprising from about 2 to about 30 weight percent fatty acids calculated as triglycerides are encompassed by the present invention. Other vitamins, and particularly fat-soluble vitamins such as vitamin A, D, E and L-carnitine optionally can be included. Where desired, a preservative such as a tocopherol can be added, typically at about 0.1% by weight.

The following examples are presented by way of illustration, not of limitation.

EXAMPLES

Example 1

The Identity of ORFs Derived from *Vibrio marinus*

Using polymerase chain reaction (PCR) with primers based on ORF 6 of *Shewanella* (Sp ORF 6) sequences (FW 5' primers CUACUACUACUACCAAGCT AAAGCACTTAACCGTG, and CUACUACUACUACAGCGAAATGCTTATCAAG for *Vibrio* and SS9 respectively and 3' BW primers: CAUCAUCAUGCGACC

AAAACCAAATGAGCTAATAC for both *Vibrio* and SS9) and genomic DNAs templates from *Vibrio* and a borophyllic *photobacter* producing EPA (provided by Dr. Bartlett, UC San Diego), resulted in PCR products of *ca.*400 bases for *Vibrio marinus* (*Vibrio*) and *ca.*900 bases for SS9 presenting more than 75% homology with corresponding fragments of Sp ORF 6 (*see* Figure 25) as determined by direct counting of homologous amino acids.

A *Vibrio* cosmid library was then prepared and using the *Vibrio* ORF 6 PCR product as a probe (*see* Figure 26); clones containing at least ORF 6 were selected by colony hybridization.

Through additional sequences of the selected cosmids such as cosmid #9 and cosmid #21, a *Vibrio* cluster (Figure 5) with ORFs homologous to, and organized in the same sequential order (ORFs 6-9) as ORFs 6-9 of *Shewanella*, was obtained (Figure 7). The *Vibrio* ORFs from this sequence are found at 17394 to 36115 and comprehend ORFs 6-9.

Table
Vibrio operon figures

17394 to 25349	length = 7956 nt
25509 to 28157	length = 2649 nt
28209 to 34262	length = 6054 nt
34454 to 36115	length = 1662 nt

The ORF designations for the *Shewanella* genes are based on those disclosed in Figure 4, and differ from those published for the *Shewanella* cluster (Yazawa *et al*, USPN 5,683,898). For instance, ORF 3 of Figure 4 is read in the opposite direction from the other ORFs and is not disclosed in Yazawa *et al* USPN 5,683,898 (See Fig. 24) for comparison with Yazawa *et al* USPN 5,683,898).

Sequences homologous to ORF 3, were not found in the proximity of ORF 6 (17000 bases upstream of ORF 6) or of ORF 9 (*ca.*4000 bases downstream of ORF 9).

Motifs characteristic of phosphopantethenyl transferases (Lambalot *et al* (1996) *Current Biology* 3:923-936) were absent from the *Vibrio* sequences screened for these motifs. In addition, there was no match to Sp ORF 3 derived probes in genomic digests of *Vibrio* and of SC2A *Shewanella* (another bacterium provided by the University of San Diego and

also capable of producing EPA). Although ORF 3 may exist in *Vibrio*, its DNA may not be homologous to that of *Sp* ORF 3 and/or could be located in portions of the genome that were not sequenced.

Figure 6 provides the sequence of an approximately 19 kb *Vibrio* clone comprising ORFs 6-9. Figures 7 and 8 compare the gene cluster organizations of the PKS-like systems of *Vibrio marinus* and *Shewanella putrefaciens*. Figures 9 through 12 show the levels of sequence homology between the corresponding ORFs 6, 7, 8 and 9, respectively.

Example 2

ORF 8 Directs DHA Production

As described in example 1, DNA homologous to *Sp* ORF 6 was found in an unrelated species, SS9 *Photobacter*, which also is capable of producing EPA. Additionally, ORFs homologous to *Sp* ORF 6-9 were found in the DHA producing *Vibrio marinus* (*Vibrio*). From these ORFs a series of experiments was designed in which deletions in each of *Sp* ORFs 6-9 that suppressed EPA synthesis in *E. coli* (Yazawa (1996) *supra*) were complemented by the corresponding homologous genes from *Vibrio*.

The *Sp* EPA cluster was used to determine if any of the *Vibrio* ORFs 6-9 was responsible for the production of DHA. Deletion mutants provided for each of the *Sp* ORFs are EPA and DHA null. Each deletion was then complemented by the corresponding *Vibrio* ORF expressed behind a *lac* promoter (Figure 13).

The complementation of a *Sp* ORF 6 deletion by a *Vibrio* ORF 6 reestablished the production of EPA. Similar results were obtained by complementing the *Sp* ORF 7 and ORF 9 deletions. By contrast, the complementation of a *Sp* ORF 8 deletion resulted in the production of C22:6. *Vibrio* ORF 8 therefore appears to be a key element in the synthesis of DHA. Figures 14 and 15 show chromatograms of fatty acid profiles from the respective complementations of *Sp* del ORF 6 with *Vibrio* ORF 6 (EPA and no DHA) and *Sp* del ORF 8 with *Vibrio* ORF 8 (DHA). Figure 16 shows the fatty acid percentages for the ORF 8 complementation, again demonstrating that ORF 8 is responsible for DHA production.

These data show that polyketide-like synthesis genes with related or similar ORFs can be combined and expressed in a heterologous system and used to produce a distinct PUFA species in the host system, and that ORF 8 has a role in determining the ultimate chain length. The *Vibrio* ORFs 6, 7, 8, and 9 reestablish EPA synthesis. In the case of

Vibrio ORF 8, DHA is also present (ca. 0.7%) along with EPA (ca. 0.6%) indicating that this gene plays a significant role in directing synthesis of DHA vs EPA for these systems.

Example 3

Requirements for Production of DHA

To determine how *Vibrio* ORFs of the cluster ORF 6-9 are used in combination with *Vibrio* ORF 8, some combinations of *Vibrio* ORF 8 with some or all of the other *Vibrio* ORFS 6-9 cluster were created to explain the synthesis of DHA.

Vibrio ORFs 6-9 were complemented with *Sp* ORF 3. The results of this complementation are presented in Figures 16b and 16c. The significant amounts of DHA measured (greater than about 9%) and the absence of EPA suggest that no ORFs other than those of *Vibrio* ORFs 6-9 are required for DHA synthesis when combined with *Sp* ORF 3. This suggests that *Sp* ORF 3 plays a general function in the synthesis of bacterial PUFAs.

With respect to the DHA vs EPA production, it may be necessary to combine *Vibrio* ORF 8 with other *Vibrio* ORFs of the 6-9 cluster in order to specifically produce DHA. The roles of *Vibrio* ORF 9 and each of the combinations of *Vibrio* ORFs (6,8), (7, 8), (8, 9), etc in the synthesis of DHA are being studied.

Example 4

Plant Expression Constructs

A cloning vector with very few restriction sites was designed to facilitate the cloning of large fragments and their subsequent manipulation. An adapter was assembled by annealing oligonucleotides with the sequences AAGCCCGGGCTT and GTACAAGCCCGGGCTTAGCT. This adapter was ligated to the vector pBluescript II SK+ (Stratagene) after digestion of the vector with the restriction endonucleases *Asp*718 and *Sst*I. The resulting vector, pCGN7769 had a single *Srf*I (and embedded *Sma*I) cloning site for the cloning of blunt ended DNA fragments.

A plasmid containing the napin cassette from pCGN3223, (USPN 5,639,790) was modified to make it more useful for cloning large DNA fragments containing multiple restriction sites, and to allow the cloning of multiple napin fusion genes into plant binary transformation vectors. An adapter comprised of the self annealed oligonucleotide of sequence CGCGATTAAATGGCGCGCCCTGCAGGCGGCCGCTGCAGGGCGC

GCCATTAAAT was ligated into the vector pBC SK+ (Stratagene) after digestion of the vector with the restriction endonuclease *Bss*HII to construct vector pCGN7765. Plasmids pCGN3223 and pCGN7765 were digested with *Nor*I and ligated together. The resultant vector, pCGN7770 (Figure 17), contains the pCGN7765 backbone and the napin seed specific expression cassette from pCGN3223.

Shewanella constructs

Genes encoding the *Shewanella* proteins were mutagenized to introduce suitable cloning sites 5' and 3' ORFs using PCR. The template for the PCR reactions was DNA of the cosmid pEPA (Yazawa *et al*, *supra*). PCR reactions were performed using Pfu DNA polymerase according to the manufacturers' protocols. The PCR products were cloned into *Srf*I digested pCGN7769. The primers CTGCAGCTCGAGACAATGTTGATT TCCTTATACTTCTGTCC and GGATCCAGATCTCTAGCTAGTCTTAGCTGAAGC TCGA were used to amplify ORF 3, and to generate plasmid pCGN8520. The primers TCTAGACTCGAGACAATGAGCCAGACCTCTAAACCTACA and CCCGGGCTC GAGCTAATTCGCCTCACTGTCGTTTGCT were used to amplify ORF 6, and generate plasmid pCGN7776. The primers GAATTCCTCGAGACAATGCCGCTGCGCATCG CACTTATC and GGTACCAGATCTTTAGACTTCCCCTTGAAGTAAATGG were used to amplify ORF 7, and generate plasmid pCGN7771. The primers GAATTCGTCG ACACAATGTCATTACCAGACAATGCTTCT and TCTAGAGTCGACTTATAC AGATTCTTCGATGCTGATAG were used to amplify ORF 8, and generate plasmid pCGN7775. The primers GAATTCGTCGACACAATGAATCCTACAGCAA CTAACGAA and TCTAGAGGATCCTTAGGCCATTCTTTGGTTTGGCTTC were used to amplify ORF 9, and generate plasmid pCGN7773.

The integrity of the PCR products was verified by DNA sequencing of the inserts of pCGN7771, pCGN8520, and pCGN7773. ORF 6 and ORF 8 were quite large in size. In order to avoid sequencing the entire clones, the center portions of the ORFs were replaced with restriction fragments of pEPA. The 6.6 kilobase *Pac*I/*Bam*HI fragment of pEPA containing the central portion of ORF 6 was ligated into *Pac*I/*Bam*HI digested pCGN7776 to yield pCGN7776B4. The 4.4 kilobase *Bam*HI/*Bg*II fragment of pEPA containing the central portion of ORF 8 was ligated into *Bam*HI/*Bg*II digested pCGN7775 to yield pCGN7775A. The regions flanking the pEPA fragment and the cloning junctions were verified by DNA sequencing.

Plasmid pCGN7771 was cut with *Xho*I and *Bgl*II and ligated to pCGN7770 after digestion with *Sa*I and *Bgl*II. The resultant napin/ORF 7 gene fusion plasmid was designated pCGN7783. Plasmid pCGN8520 was cut with *Xho*I and *Bgl*II and ligated to pCGN7770 after digestion with *Sa*I and *Bgl*II. The resultant napin/ORF 3 gene fusion
5 plasmid was designated pCGN8528. Plasmid pCGN7773 was cut with *Sa*I and *Bam*HI and ligated to pCGN7770 after digestion with *Sa*I and *Bgl*II. The resultant napin/ORF 9 gene fusion plasmid was designated pCGN7785. Plasmid pCGN7775A was cut with *Sa*I and ligated to pCGN7770 after digestion with *Sa*I. The resultant napin/ORF 8 gene fusion plasmid was designated pCGN7782. Plasmid pCGN7776B4 was cut with *Xho*I
10 and ligated to pCGN7770 after digestion with *Sa*I. The resultant napin/ORF 6 gene fusion plasmid was designated pCGN7786B4.

A binary vector for plant transformation, pCGN5139, was constructed from pCGN1558 (McBride and Summerfelt (1990) *Plant Molecular Biology*, 14:269-276). The polylinker of pCGN1558 was replaced as a *Hind*III/*Asp*718 fragment with a
15 polylinker containing unique restriction endonuclease sites, *As*CI, *Pac*I, *Xba*I, *Swa*I, *Bam*HI, and *Not*I. The *Asp*718 and *Hind*III restriction endonuclease sites are retained in pCGN5139. pCGN5139 was digested with *Not*I and ligated with *Not*I digested pCGN7786B4. The resultant binary vector containing the napin/ORF 6 gene fusion was designated pCGN8533. Plasmid pCGN8533 was digested with *Sse*8387I and ligated with
20 *Sse*8387I digested pCGN7782. The resultant binary vector containing the napin/ORF 6 gene fusion and the napin/ORF 8 gene fusion was designated pCGN8535 (Figure 18).

The plant binary transformation vector, pCGN5139, was digested with *Asp*718 and ligated with *Asp*718 digested pCGN8528. The resultant binary vector containing the napin/ORF 3 gene fusion was designated pCGN8532. Plasmid pCGN8532 was digested
25 with *Not*I and ligated with *Not*I digested pCGN7783. The resultant binary vector containing the napin/ORF 3 gene fusion and the napin/ORF 7 gene fusion was designated pCGN8534. Plasmid pCGN8534 was digested with *Sse*8387I and ligated with *Sse*8387I digested pCGN7785. The resultant binary vector containing the napin/ORF 3 gene fusion, the napin/ORF 7 gene fusion and the napin/ORF 9 gene fusion was designated
30 pCGN8537 (Figure 19).

Vibrio constructs

The *Vibrio* ORFs for plant expression were all obtained using *Vibrio* cosmid #9 as a starting molecule. *Vibrio* cosmid #9 was one of the cosmids isolated from the *Vibrio* cosmid library using the *Vibrio* ORF 6 PCR product described in Example 1.

5 A gene encoding *Vibrio* ORF 7 (Figure 6) was mutagenized to introduce a *SalI* site upstream of the open reading frame and *Bam*HI site downstream of the open reading frame using the PCR primers: TCTAGAGTCGACACAATGGCGGAATTAGCTG TTATTGGT and GTCGACGGATCCCCTATTTGTTTCGTGTTTGCTATATG. A gene encoding *Vibrio* ORF 9 (Figure 6) was mutagenized to introduce a *Bam*HI site upstream of the open reading frame and an *Xho*HI site downstream of the open reading frame using the PCR primers: GTCGACGGATCCACAATGAATATAGTAAGTAATCATTCGGCA and GTCGACCTCGAGTTAATCACTCGTACGATAACTTGCC. The restriction sites were introduced using PCR, and the integrity of the mutagenized plasmids was verified by DNA sequence. The *Vibrio* ORF 7 gene was cloned as a *SalI*-*Bam*HI fragment into the napin cassette of *Sal*-*Bgl*II digested pCGN7770 (Figure 17) to yield pCGN8539. The *Vibrio* ORF 9 gene was cloned as a *SalI*-*Bam*HI fragment into the napin cassette of *Sal*-*Bal*II digested pCGN7770 (Figure 17) to yield pCGN8543.

Genes encoding the *Vibrio* ORF 6 and ORF 8 were mutagenized to introduce *SalI* sites flanking the open reading frames. The *SalI* sites flanking ORF 6 were introduced using PCR. The primers used were: CCCGGGTCGACACAATGGCTAAAAAGAACA CCACATCGA and CCCGGGTCGACTCATGACATATCGTTCAAAATGTCACTGA. The central 7.3 kb *Bam*HI-*Xho*I fragment of the PCR product was replaced with the corresponding fragment from *Vibrio* cosmid #9. The mutagenized ORF 6 were cloned into the *SalI* site of the napin cassette of pCGN7770 to yield plasmid pCGN8554.

25 The mutagenesis of ORF 8 used a different strategy. A *Bam*HI fragment containing ORF 8 was subcloned into plasmid pH79 to yield cosmid #9". A *SalI* site upstream of the coding region was introduced on an adapter comprised of the oligonucleotides TCGACATGGAAAATATTGCAGTAGTAGGTATTGCTAATTT GTTC and CCGGGAACAAATTAGCAATACCTACTACTGCAATATTTTCCATG. The adapter was ligated to cosmid #9" after digestion with *SalI* and *Xma*I. A *SalI* site was introduced downstream of the stop codon by using PCR for mutagenesis. A DNA fragment containing the stop codon was generated using cosmid #9" as a template with the primers TCAGATGAACTTTATCGATAC and TCATGAGACGTCGTCGACTTA

CGCTTCAACAATACT. The PCR product was digested with the restriction endonucleases *Cla*I and *Aat*II and was cloned into the cosmid 9" derivative digested with the same enzymes to yield plasmid 8P3. The *Sa*II fragment from 8P3 was cloned into *Sa*II digested pCGN7770 to yield pCGN8515.

5 pCGN8532, a binary plant transformation vector that contains a *Shewanella* ORF 3 under control of the napin promoter was digested with *Not*I, and a *Not*I fragment of pCGN8539 containing a napin *Vibrio* ORF 7 gene fusion was inserted to yield pCGN8552. Plasmid pCGN8556 (Figure 23), which contains *Shewanella* ORF 3, and *Vibrio* ORFs 7 and 9 under control of the napin promoter was constructed by cloning the
10 *Sse*8357 fragment from pCGN8543 into *Sse*8387 digested pCGN8552.

 The *Not*I digested napin/ORF 8 gene from plasmid pCGN8515 was cloned into a *Not*I digested plant binary transformation vector pCGN5139 to yield pCGN8548. The *Sse*8387 digested napin/ORF 6 gene from pCGN8554 was subsequently cloned into the *Sse*8387 site of pCGN8566. The resultant binary vector containing the napin/ORF 6 gene
15 fusion and napin/ORF 8 gene fusion was designated pCGN8560 (Figure 22).

Example 5

Plant Transformation and PUFA Production

EPA production

20 The *Shewanella* constructs pCGN8535 and pCGN8537 can be transformed into the same or separate plants. If separate plants are used, the transgenic plants can be crossed resulting in heterozygous seed which contains both constructs.

 pCGN8535 and pCGN8537 are separately transformed into *Brassica napus*. Plants are selected on media containing kanamycin and transformation by full length
25 inserts of the constructs is verified by Southern analysis. Immature seeds also can be tested for protein expression of the enzyme encoded by ORFs 3, 6, 7, 8, or 9 using western analysis, in which case, the best expressing pCGN8535 and pCGN8537 T₁
transformed plants are chosen and are grown out for further experimentation and crossing. Alternatively, the T₁ transformed plants showing insertion by Southern are crossed to one
30 another producing T₂ seed which has both insertions. In this seed, half seeds may be analyzed directly from expression of EPA in the fatty acid fraction. Remaining half-seed

of events with the best EPA production are grown out and developed through conventional breeding techniques to provide *Brassica* lines for production of EPA.

Plasmids pCGN7792 and pCGN7795 also are simultaneously introduced into *Brassica napus* host cells. A standard transformation protocol is used (see for example
5 USPN 5,463,174 and USPN 5,750,871, however *Agrobacteria* containing both plasmids are mixed together and incubated with *Brassica* cotyledons during the cocultivation step. Many of the resultant plants are transformed with both plasmids.

DHA production

10 A plant is transformed for production of DHA by introducing pCGN8556 and pCGN8560, either into separate plants or simultaneously into the same plants as described for EPA production.

Alternatively, the *Shewanella* ORFs can be used in a concerted fashion with ORFs 6 and 8 of *Vibrio*, such as by transforming with a plant the constructs pCGN8560 and
15 pCGN7795, allowing expression of the corresponding ORFs in a plant cell. This combination provides a PKS-like gene arrangement comprising ORFs 3, 7 and 9 of *Shewanella*, with an ORF 6 derived from *Vibrio* and also an OFR 8 derived from *Vibrio*. As described above, ORF 8 is the PKS-like gene which controls the identity of the final PUFA product. Thus, the resulting transformed plants produce DHA in plant oil.

Example 6

Transgenic plants containing the *Shewanella* PUFA genes

Brassica plants

Fifty-two plants cotransformed with plasmids pCGN8535 and pCGN8537 were
25 analyzed using PCR to determine if the *Shewanella* ORFs were present in the transgenic plants. Forty-one plants contained plasmid pCGN8537, and thirty-five plants contained pCGN8535. 11 of the plants contained all five ORFs required for the synthesis of EPA. Several plants contained genes from both of the binary plasmids but appeared to be missing at least one of the ORFs. Analysis is currently being performed on approximately
30 twenty additional plants.

Twenty-three plants transformed with pCGN8535 alone were analyzed using PCR to determine if the *Shewanella* ORFs were present in the transgenic plants. Thirteen of

these plants contained both *Shewanella* ORF 6 and *Shewanella* ORF 8. Six of the plants contained only one ORF.

Nineteen plants transformed with pCGN8537 were alone analyzed using PCR to determine if the *Shewanella* ORFs were present in the transgenic plants. Eighteen of the plants contained *Shewanella* ORF 3, *Shewanella* ORF 7, and *Shewanella* ORF 9. One plant contained *Shewanella* ORFs 3 and 7.

Arabidopsis

More than 40 transgenic Arabidopsis plants cotransformed with plasmids pCGN8535 and pCGN8537 are growing in our growth chambers. PCR analysis to determine which of the ORFs are present in the plants is currently underway.

By the present invention PKS-like genes from various organisms can now be used to transform plant cells and modify the fatty acid compositions of plant cell membranes or plant seed oils through the biosynthesis of PUFAs in the transformed plant cells. Due to the nature of the PKS-like systems, fatty acid end-products produced in the plant cells can be selected or designed to contain a number of specific chemical structures. For example, the fatty acids can comprise the following variants: Variations in the numbers of keto or hydroxyl groups at various positions along the carbon chain; variations in the numbers and types (*cis* or *trans*) of double bonds; variations in the numbers and types of branches off of the linear carbon chain (methyl, ethyl, or longer branched moieties); and variations in saturated carbons. In addition, the particular length of the end-product fatty acid can be controlled by the particular PKS-like genes utilized.

All publications and patent applications mentioned in this specification are indicative of the level of skill of those skilled in the art to which this invention pertains. All publications and patent applications are herein incorporated by reference to the same extent as if each individual publication or patent application was specifically and individually indicated to be incorporated by reference.

The invention now being fully described, it will be apparent to one of ordinary skill in the art that many changes and modifications can be made thereto without departing from the spirit or scope of the appended claims.

What is claimed is:

1. An isolated nucleic acid comprising:
a *Vibrio marinus* nucleotide sequence selected from the group consisting of the ORF 6, ORF 7, ORF 8 and ORF 9 as shown in Figure 6.
2. An isolated nucleic acid comprising:
a nucleotide sequence which encodes a polypeptide of a polyketide-like synthesis system, wherein said system produces a docosahexenoic acid when expressed in a host cell.
3. The isolated nucleic acid according to Claim 2, wherein said nucleotide sequence is derived from a marine bacterium.
4. The isolated nucleic acid according to Claim 2, wherein said nucleotide sequence is a *Vibrio marinus* ORF 8 as shown in Figure 6.
5. An isolated nucleic acid comprising:
a nucleotide sequence which is substantially identical to a sequence of at least 50 nucleotides of a *Vibrio marinus* nucleotide sequence selected from the group consisting of ORF 6, ORF 7, ORF 8 and ORF 9 as shown in Figure 6.
6. A recombinant microbial cell comprising at least one copy of an isolated nucleic acid according to Claim 1 or Claim 2.
7. The recombinant microbial cell according to Claim 6, wherein said cell comprises each element of a polyketide-like synthesis system required to produce a long chain polyunsaturated fatty acid.
8. The recombinant microbial cell according to Claim 7, wherein said cell is a eukaryotic cell.
9. The recombinant microbial cell according to Claim 8, wherein said eukaryotic cell is a fungal cell, an algae cell or an animal cell.

10. The recombinant microbial cell according to Claim 9, wherein said fungal cell is a yeast cell and said algae cell is a marine algae cell.

11. The recombinant microbial cell according to Claim 6, wherein said cell is a
5 prokaryotic cell.

12. The recombinant microbial cell according to Claim 11, wherein said cell is a bacterial cell or a cyanobacterial cell.

10 13. The microbial cell according to Claim 6, wherein said recombinant microbial cell is enriched for 22:6 fatty acids as compared to a non-recombinant microbial cell which is devoid of said isolated nucleic acid.

14. A method for production of docosahexenoic acid in a microbial cell culture,
15 said method comprising:

growing a microbial cell culture having a plurality of microbial cells, wherein said microbial cells or ancestors of said microbial cells were transformed with a vector comprising one or more nucleic acids having a nucleotide sequence which encodes a polypeptide of a polyketide synthesizing system, wherein said one or more nucleic acids
20 are operably linked to a promoter, under conditions whereby said one or more nucleic acids are expressed and docosahexenoic acid is produced in said microbial cell culture.

15. A method for production of a long chain polyunsaturated fatty acid in a plant cell, said method comprising:

25 growing a plant having a plurality of plant cells, wherein said plant cells or ancestors of said plant cells were transformed with a vector comprising one or more nucleic acids having a nucleotide sequence which encodes one or more polypeptides of a polyketide synthesizing system which produces a long chain polyunsaturated fatty acid, wherein each of said nucleic acids are operably linked to a promoter functional in a plant
30 cell, under conditions whereby said polypeptides are expressed and a long chain polyunsaturated fatty acid is produced in said plant cells.

16. The method according to Claim 15, wherein said long chain polyunsaturated fatty acid produced in said plant cells is a 20:5 and 22:6 fatty acid.

17. The method according to Claim 15, wherein said nucleic acids comprise nucleotide sequences encoding any one of the polypeptides selected from the group consisting of *Vibrio marinus* ORF 6, ORF 7, ORF 8 and ORF 9 as shown in Figure 6 and *Shewanella putrefaciens* ORF 3, ORF 6, ORF 7, ORF 8 and ORF 9 as shown in Figure 4.

18. The method according to Claim 15, wherein said nucleic acid constructs are derived from two or more polyketide synthesizing systems.

19. A recombinant plant cell which produces an long chain polyunsaturated fatty acid exogenous to said plant cell, wherein said recombinant plant cell is produced according to a method comprising:

transforming a plant cell or an ancestor or said plant cell with a vector comprising one or more nucleic acids having a nucleotide sequence which encodes one or more polypeptides of a polyketide synthesizing system which produces a long chain polyunsaturated fatty acid, wherein each of said nucleic acids are operably linked to a promoter functional in said plant cell whereby a recombinant plant cell is obtained; and growing said recombinant plant cell under conditions whereby said polypeptides are expressed and a long chain polyunsaturated fatty acid is produced in said plant cell.

20. The recombinant plant cell according to Claim 19, wherein said recombinant plant cell is a recombinant seed cell.

21. The recombinant plant cell according to Claim 20, wherein said recombinant seed cell is a recombinant embryo cell.

22. The method according to Claim 15, wherein said long chain polyunsaturated fatty acid produced in said plant cells is eicosapentenoic acid.

23. The method according to Claim 15, wherein said long chain polyunsaturated fatty acid produced in said plant cells is docosahexenoic acid.

24. The recombinant plant cell according to Claim 19, wherein said recombinant plant cell is from a plant selected from the group consisting of *Brassica*, soybean, safflower, and sunflower.

5 25. A plant oil produced by a recombinant plant cell according to Claim 19, wherein said plant oil comprises eicosapentenoic acid.

26. A plant oil produced by a recombinant plant cell according to Claim 19, wherein said plant oil comprises docosahexenoic acid.

10

27. The plant oil according to Claim 25 or Claim 26, wherein said plant oil is encapsulated.

28. A dietary supplement comprising a plant oil according to Claim 27.

15

29. A recombinant *E. coli* cell which produces docosahexenoic acid.

30. A plant oil comprising eicosapentenoic acid.

31. A plant oil comprising docosahexenoic acid.

20

32. The recombinant microbial cell according to Claim 12, wherein said bacterial cell is a lactobacillus cell.

Fig. 1 Organization of Shewanella EPA Genes and Clones Obtained from the Sagami Chemical Institute.

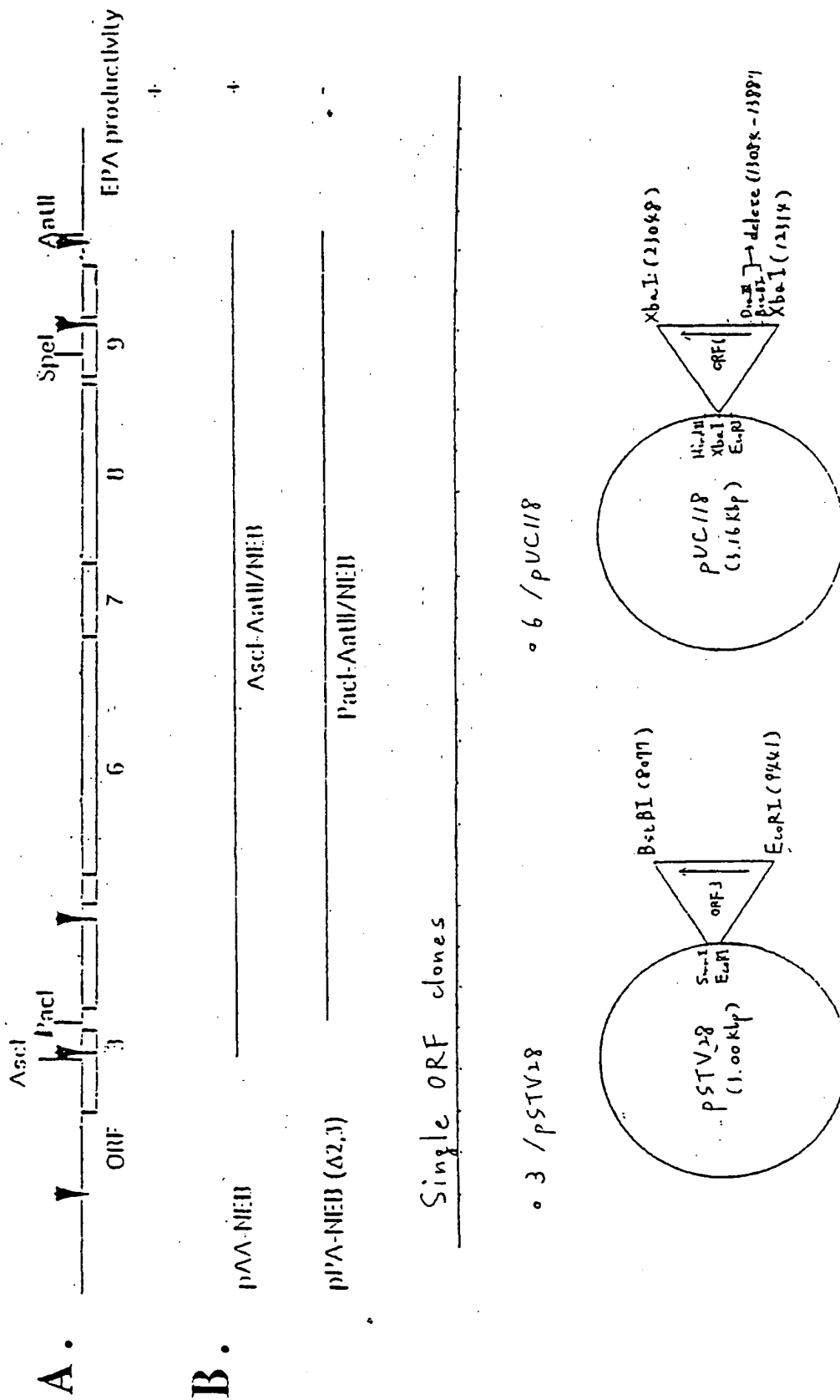
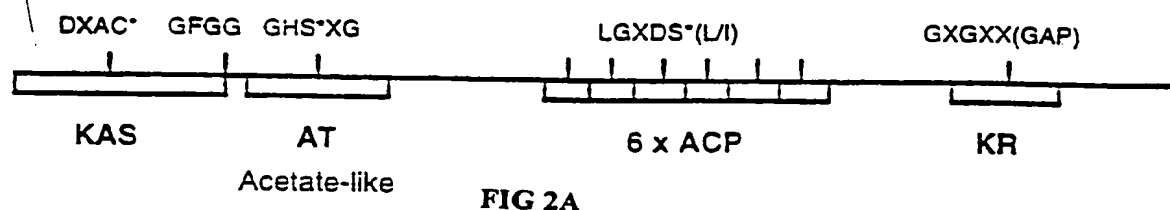
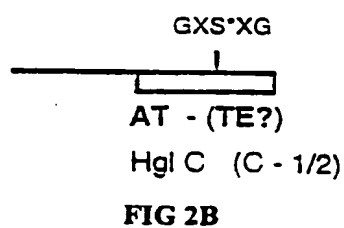
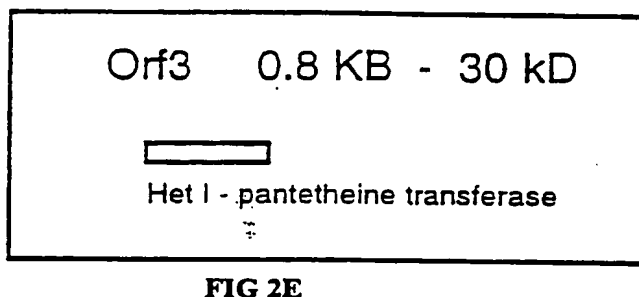
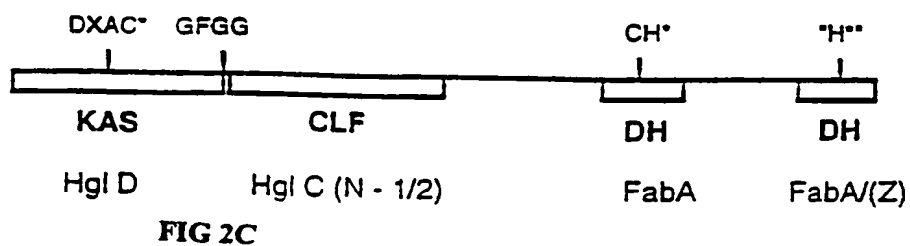
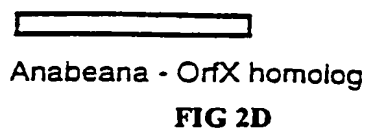
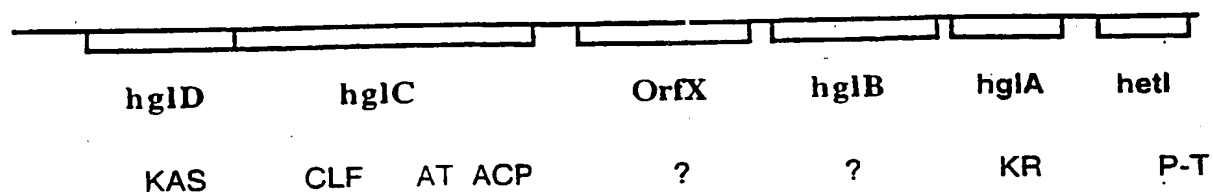
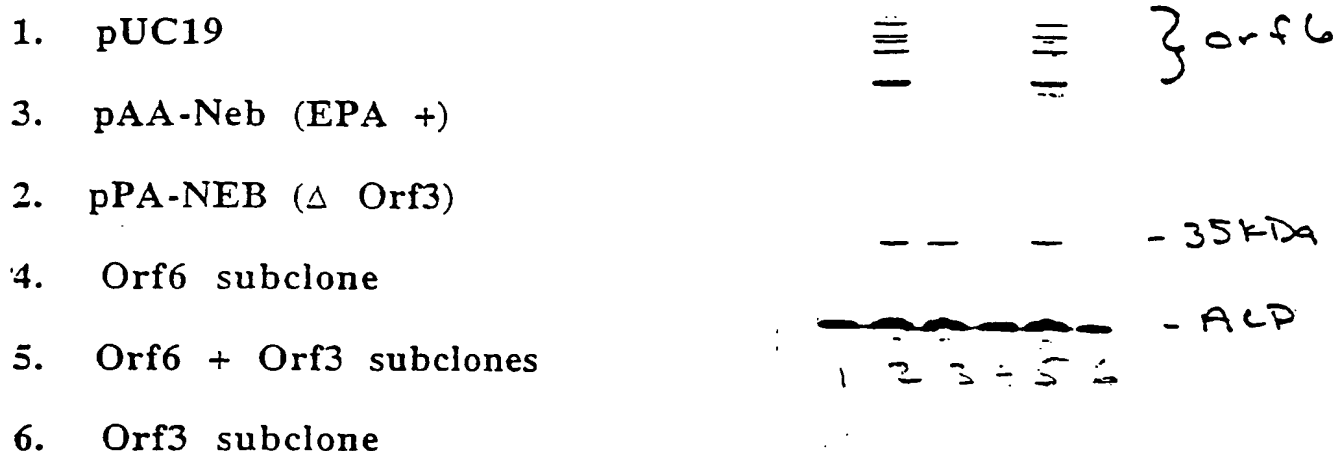


Fig. 2**SYNECHOCYSTIS PALLIDUS ORFs****Motifs - Domains - Homologies****Orf6 8.3 KB - 293 kD****Orf7 2.3 KB - 84 kD****Orf3 0.8 KB - 30 kD****Orf8 6.0 KB - 217 kD****Orf9 1.6 KB - 59 kD**



**Anabeana "PKS" Genes Involved in Heterocyst
Glycolipid Synthesis****

Fig 3. Orf3 Encodes a Phosphopantetheine Transferase



Autoradiograph of [C14] β -Alanine labelled proteins from *E. coli* (strain SJ16) cells transformed with the above listed plasmids. Cells were grown in the presence of [C14] β -alanine and the appropriate antibiotics. Proteins were extracted, separated by SDS-PAGE and transferred to a PVDF membrane prior to autoradiography. ACP and an unknown (but previously observed) 35 kD protein were labelled in all of the samples. The high molecular mass proteins detected in lanes 2 and 5 are full-length (largest band) and truncated products of the *Shewanella* Orf6 gene (confirmed by Western analysis - data not shown). *E. coli* strain SJ16 is conditionally blocked in β -alanine synthesis.

Sequence Range: 1 to 37895

20 40 60 80
GATCTCTTAC AAAGAACTA TCTCAATGTG AATTTAACTT TAATTCCGTT TAATTACGGC CTGATAGAGC ATCACCCAAT
100 120 140 160
CAGCCATAAA ACTGTAAAGT GGGTACTCAA AGGTGGCTGG GCGATTCTTC TCAAATACAA AGTGCCCAAC CCAAGCAAAT
180 200 220 240
CCATATCCGA TAACAGGTAA AAGTAGCAAT AAACCCAGC GCTGAGTTAG TAATACATAA GCGAATAATA GGATCACTAA
260 280 300 320
ACTACTGCCG AAATAGTGTA ATATTCGACA GTTCTATGCT TGATGTTGAG ATAAATAAAA ACGGTAAAAT TCAGCAAAAG
340 360 380 400
AACGATAGCG CTTACTCATT ACTCACACCT CGGTAAAAA GCAACTCGCC ATTAACCTGG CCAATCGTCA GTTGTCTTAT
420 440 460 480
CGTCTCAAAG TTATGCCGAC TAAATAACTC TATATGTGCA TTATGATTAG CAAAACTCC GATACCATCA AGATGAAGTT
500 520 540 560
GTTTCATACA CCAACTCAA ACTGCGTCGA TAAGCTTACT GCCATAGCCC TTGCCTTGCT CCACATTGCG GATAGCAATA
580 600 620 640
AACTGTAAAA TGCCACATTG GCCACTTGGT AAGCTCTCTA TAATCTGATT TTCTTTGTTA ATAAGTGCCT GAGTGAATA
660 680 700 720
CCAACCAGTA CTTAACAACA TCTTTAAACG CCAATGCCAA AAACCGGCTT CACCTAAGGG AACCTGCTGA GTCACTATGC
740 760 780 800
AGGCTACGCC TATCAATCTA TCCCCAACGA ACATACCAAT AAGTGCTTGC TCCTGTTGCC AGAGCTCATT GAGTCTCTCT
820 840 860 880
CGAATAGCCC CGCGAAGCTT TTGCTCATAC TGGCCTTGAT CACCACTAAA AAGTGTTTCG ATAAAAAAGG GATCATCATG
900 920 940 960
ATAGGCGTTA TAGAGAATAG AGGCTGCTAT GCGTAAATCT TCTGCCGTGA GATAAACTGC ACGACACTCT TCCATGGCTT
980 1000 1020 1040
GATCTTCCAT TGTATTSTC CTTGACCTTG ATCACACAAC ACCAATGTAA CAAGACTGTA TAGAAGTGCA ATTAATAATC
1060 1080 1100 1120
AATTCGTGCA TTAAGCAGST CAGCATTCTT TTGCTAAACA AGCTTTATTG GCTTTGACAA AACTTTGCCT AGACTTTAAC
1140 1160 1180 1200
GATAGAAATC ATAATGAAAG AGAAAAGCTA CAACCTAGAG GGGAAATAATC AAACAACTGC TAAGATCTAG ATAATGTAAT
1220 1240 1260 1280
AAACACCGAG TTTATCGACC ATACTTAGAT AGAGTCATAG CAACGAGAAT AGTTATGGAT ACAACGCCGC AAGATCTATC
1300 1320 1340 1360
ACACCTGTTT TTACAGCTAG GATTAGCAAA TGATCAACCC GCAATTGAAC AGTTTATCAA TGACCATCAA TTAGCGGACA
1380 1400 1420 1440
ATATATTGCT ACATCAAGCA AGCTTTTGGA GCCCATCGCA AAAGCACTTC TTAATTGAGT CATTTAATGA AGATGCCGAG
1460 1480 1500 1520
TGGACCGAAG TCATCGACCA CTTAGACACC TTATTAAGAA AAACTAACC ATTACAACAG CAACTTTAAA TTTTGGCGTA
1540 1560 1580 1600
AGCCATCTCC CCCCACCCCA CAACAGCGTT GTTGCTTATG ACCACTGGAG TACATTGCTC TTTAGTCGTT TTACCATCAC
1620 1640 1660 1680
CATGGGTACG TTGAGTCCGA TAAAAAGCA CATAAACTTC TTTATCGGCC TGAATATAGG CTCGTTAAA ATCAGCTGTT
1700 1720 1740 1760
CCCATTAAG TAACCACTTG CTCTTACTC ATGCCTAGAG ATATCTTTGT CAAATTGTCA CGGTTTTTAT CTTGAGTTTT

Fig. 4
1/30

1780 1800 1820 1840
CTCCCAAGCA CCGTGATTAT CCCAGTCAGA TTCCCATCA CCAACATTGA CCACACAGCC CGTTAGCCCT AAGCTTGCAA
1860 1880 1900 1920
TCCCAAAACA TGCTAAACCT AATAATTAT TTTTCATTT AACTTCCGT TATGACATTA TTTTGTCTTA GAAGAAAAGC
1940 1960 1980 2000
AACTTACATG CCAAAACACA AGCTGTTGTT TTAAATGACT TTATTTATTA TTAGCCTTTT AGGATATGCC TAGAGCAATA
2020 2040 2060 2080
ATAATTACCA ATGTTTAAGG AATTGACTA ACTATGAGTC CGATTGAGCA AGTGCTAACA GCTGCTAAAA AAATCAATGA
2100 2120 2140 2160
ACAAGGTAGA GAACCAACAT TAGCATTGAT TAAACCAAAA CTTGGTAATA GCATCCCAAT GCGCGAGTTA ATCCAAGGTT
2180 2200 2220 2240
TGCAACAGTT TAAGTCTATG AGTGCAAG AAAGACAAGC AATACCTAGC AGCTTAGCAA CAGCAAAAGA AACTCAATAT
2260 2280 2300 2320
GGTCAATCAA GCTTATCTCA ATCTGAACAA GCTGATAGGA TCCTCCAGCT AGAAAACGCC CTCAATGAAT TAAGAAAACGA
2340 2360 2380 2400
ATTTAATGGG CTAAAAAGTC AATTGATAA CTTACAACAA AACCTGATGA ATAAAGAGCC TGACACCAA TGCATGTAAT
2420 2440 2460 2480
TGAACACGA TTTGAATGTT TTGATAACAC CACGATTACT GCAGCAGAAA AAGCCATTAA TGGTTTGCTT GAAGCTTATC
2500 2520 2540 2560
GAGCCAATGG CCAGGTTCTA GGTGCTGAAT TTGCCGTTGC ATTTAACGAT GGTGAGTTTA AAGCACGCAT GTTAACCCCA
2580 2600 2620 2640
GAAAAAGCA GCTTATCTAA ACGCTTTAAT AGTCCTTGGG TAAATAGTGC ACTCGAAGAG CTAACCGAAG CCAAATTGCT
2660 2680 2700 2720
TGCGCCACGT GAAAAGTATA TTGGCCAAGA TATTAATTCT GAAGCATCTA GCCAAGACAC ACCAAGTTGG CAGCTACTTT
2740 2760 2780 2800
ACACAAGTTA TGTGCACATG TGCTCACCAC TAAGAAATGG CGACACCTTG CAGCCTATTC CACTGTATCA AATTCCAGCA
2820 2840 2860 2880
ACTGCCAAGC GCGATCATAA ACGAATGATC CGTTGGCAAA CAGAATGGCA AGCTTGTGAT GAATTGCAAA TGGCCGCAGC
2900 2920 2940 2960
TACTAAAGCT GAATTGCGG CACTTGAAGA GCTAACCAGT CATCAGAGTG ATCTATTTAG GCGTGGTTGG GACTTACGTG
2980 3000 3020 3040
GCAGAGTCGA ATACTTGACG AATATTCCGA CCTATTACTA TTTATACCGT GTTGGCGGTG AAAGCTTAGC AGTAGAAAAG
3060 3080 3100 3120
CAGCGCTCTT GTCCTAAGTG TGGCAGTCAA GAATGGCTGC TCGATAAACC ATTATTGGAT ATGTTCCATT TTCGCTGTGA
3140 3160 3180 3200
CACCTGCCGC ATCGTATCTA ATATCTCTTG GGACCATTTA TAACTCTTCC GAGTCTTATC AACTAGAGT TTAGTCAGCA
3220 3240 3260 3280
TAAAAATGGC GCTTATATTT CAATTAAAAG AAATATAAGC GCCATTTTCA TCGATACTAT ATATCAGCAG ACTATTTTCC
3300 3320 3340 3360
GCGTAAATTA GCCCACATTA ATTTCAATTCT TTGCCAGATC CCTGGATGAT CTAGTTGTGG CATCGACTCT TCAATAGGTT
3380 3400 3420 3440
TAACCGCAGG TGTAACCTTT GGAGTCAATT CGTTTATAAA CTCGTTTAAA CTGTCATTA ATTTAACGCT TTGTACTTCA
3460 3480 3500 3520
CCTGGAATTT CAATCCATAC GCTGCCATCA CTATTATTAA CCGTCAACAT TTTATCTTCA TCATCAAGAA TACCAATAAA

Fig. 4
2/30

3540 3560 3580 3600
CCAAGTCGGC TCTTGCTTAA GCTTTCTCTT CATCATTAAA TGACCAATGA TGTGTTTGTG TAAGTATTCA AAATCAGTTT
3620 3640 3660 3680
GATCCACAC TTGGATTAGC TCACCTTGGC CCCATTGTGA GTCAAAAAAT AGCGGTGCAG AAAAATGACT GCCAAAAAT
3700 3720 3740 3760
GGATTAATTT CTGCAGATAA TGTCATTTC AAGTGTGTTT CAACATTAGC AAATTCACCA GGTGTTTGAC GTACAACCGA
3780 3800 3820 3840
TTGCCAAAAC ACTGCGCCAT CGGAGCCCGC TTCGGCGACA ACACACTCAG ACTTTTGTCC TTGCGCATAA TATCTTGGCT
3860 3880 3900 3920
GTTACCAAG CTATCCATG TAGGCTTGT GATATTTAGA TAAAAAAGA TCTAAAGCAG GTAAAGAAGA CACTTAAGCC
3940 3960 3980 4000
AGTTCCAAA TCAGTTATAA TAGGGGTCTA TTTTGACATG GAAACCGTAT TGATGACACA ACATCATGAT CCCTACAGTA
4020 4040 4060 4080
ACGCCCCCGA ACTTCTGAA TTAACCTTAG GAAAGTCGAC CGGTTATCAA GAGCAGTATG ATGCATCTTT ACTACAAGCG
4100 4120 4140 4160
TGCCGCGTAA ATTAACCGT GATGCTATCG GTCTAACCAA TGAGCTACCT TTTTCATGGCT GTGATATTG GACTGGCTAC
4180 4200 4220 4240
GAACTGTCTT GGCTAAATGC TAAAGGCAAG CCAATGATTG CTATTGCAGA CTTTAACCTA AGTTTGTATA GTAAAAATCT
4260 4280 4300 4320
GATCGAGTCT AAGTCGTTA AGCTGTATTT AAACAGCTAT AACCAAAACAC GATTGATAG CGTTCAAGCG GTTCAAGAAC
4340 4360 4380 4400
GTTTAACTGA AACTTAAGC GCCTGTGCCC AAGGCACAGT TACGGTAAAA GTGATTGAAC CTAAGCAATT TAACCACCTG
4420 4440 4460 4480
AGAGTGGTTG ATATGCCAGG TACCTGCATT GACGATTTAG ATATTGAAGT TGATGACTAT AGCTTTAACT CTGACTATCT
4500 4520 4540 4560
CACCGACAGT GTTGATGACA AAGTCATGGT TGCTGAAACG CTAACGTCAA ACTTATTGAA ATCAAACCTG CTAATCACTT
4580 4600 4620 4640
CTCAGCCTGA CTGGGTGACA GTGATGATCC GTTATCAAGG GCCTAAGATA GACCGTGAAA AGCTACTTAG ATATCTGATT
4660 4680 4700 4720
TCATTAGAC AGCACAATGA ATTTATGAG CAGTGTGTTG AGCGTATATT TGTTGATTGA AAGCACTATT GCCAATGTGC
4740 4760 4780 4800
CAAACCTACT GTCTATGCAC GTTATACCCG CCGTGGTGGT TTAGATATCA ACCCATATCG TAGCGACTTT GAAAACCTG
4820 4840 4860 4880
CAGAAAATCA GCGCCTAGCG AGACAGTAAT TGATTGCAGT ACCTACAAAA AACAAATGCCT ATAAGCCAAG CTTATGGGCA
4900 4920 4940 4960
TTTTTATATT ATCAACTTGT CATCAAACCT CAGCCGCCAA GCCTTTTGTG TTTATCGCTA AATTAAGCCG CTCTCTCAGC
4980 5000 5020 5040
CAAATATTG CAGGATTTG CTCTAATTTA TGGCTCCACA CCATGAAATA CTCTATCGGC TCTACCGCAA AAGGTAAGTC
5060 5080 5100 5120
AAATACCTGT AAGCCAAACA GCTTGCATA TTCGTGAGT TGGGCTTTG ACGCGATAGC TAACGCATCA CTTTTTGAGG
5140 5160 5180 5200
CAACCGACAT CATACTTAAT ATTGATGATT GCTCGCTGTG CATTTGCCCT GCCGGTAACA CCTGTTTAGT CAGCAAGTCG
5220 5240 5260 5280
GCAACACTTA AATTGTAGCG GCGCATCTTA AAAATAATAT GCTTTTCATT AAAGTATTGC TCTTGCCTCA ACCCACCTTG
5300 5320 5340 5360

Fig 4
3/30

GATCCTTGGG TGAGCATTTC GTGCCACACA AACTAATTTA TCCTGCATTA CTTTTTGACT CTTAAATGCC GCAGATTCTG
5380 5400 5420 5440
GCAGCCAAAT ATCTAAGGCT AAATCCACCT TTTCTAGTTG TAGGTCCATC TGCAACTCTT CTTCAATGAG CGGCGGCTCA
5460 5480 5500 5520
CGAAATACAA TATTAATTGC AGTGCCCTGT AACACTTGCT CAATTTGATC TTGCAAGAGT TGTATTGCCG ACTCGCTGGC
5540 5560 5580 5600
ATACACATAA AAAGTTTCGCT CACTTGAAGT GGGGTCAAAT GCTTCAAAGC TAGTCGCAAC TTGCTCAATT GTTGACATAG
5620 5640 5660 5680
CGCCCCGCGAG CTGTTGATAA AGCGTCATCG CACTTGCGGT AGGTTTAACT CCCCTACCCA CTCGAGTAAA CAACTCTTCT
5700 5720 5740 5760
CCAACAATAC TTTTAGCCT CGAAATCGCA TTACTAACCG ACGACTGAGT CAAATCCAGC TCTTCTGCCG CCCGGCTAAA
5780 5800 5820 5840
AGATGAGGTG CGATACACCG CAGTAAAAAC GCGAAATAAA TTAAGATCAA AAGCTTTTTC CTGCGACATA AATCAGCTAT
5860 5880 5900 5920
CTCCTTATCC TTATCCTTAT CCTTATAAAA AGTTAGCTCC AGAGCACTCT AGCTCAAAAA CAACTCAGCG TATTAAGCCA
5940 5960 5980 6000
ATATTTTGGG AACTCAATTA ATATTCATAA TAAAAGTATT CATAATATAA ATACCAAGTC ATAATTAGC CCTAATTATT
6020 6040 6060 6080
AATCAATTCA AGTTACCTAT ACTGGCCTCA ATTAAGCAAA TGTCTCATCA GTCTCCCTGC AACTAAATGC AATATTGAGA
6100 6120 6140
CATAAAGCTT TGAAGTATT CAATCTTACG AGGTAACCT ATG AAA CAG ACT CTA ATG GCT ATC TCA ATC ATG
M K Q T L M A I S I M>
6160 6180 6200
TCG CTT TTT TCA TTC AAT GCG CTA GCA GCG CAA CAT GAA CAT GAC CAC ATC ACT GTT GAT TAC GAA
S L F S F N A L A A Q H E H D H I T V D Y E>
6220 6240 6260 6280
GGG AAA GCC GCA ACA GAA CAC ACC ATA GCT CAC AAC CAA GCT GTA GCT AAA ACA CTT AAC TTT GCC
G K A A T E H T I A H N Q A V A K T L N F A>
6300 6320 6340
GAC ACG CGT GCA TTT GAG CAA TCG TCT AAA AAT CTA GTC GCC AAG TTT GAT AAA GCA ACT GCC GAT
D T R A F E Q S S K N L V A K F D K A T A D>
6360 6380 6400
ATA TTA CGT GCC GAA TTT GCT TTT ATT AGC GAT GAA ATC CCT GAC TCG GTT AAC CCG TCT CTC TAC
I L R A E F A F I S D E I P D S V N P S L Y>
6420 6440 6460 6480
CGT CAG GCT CAG CTT AAT ATG GTG CCT AAT GGT CTG TAT AAA GTG AGC GAT GGC ATT TAC CAG GTC
R Q A Q L N M V P N G L Y K V S D G I Y Q V>
6500 6520 6540
CGC GGT ACC GAC TTA TCT AAC CTT ACA CTT ATC CGC AGT GAT AAC GGT TGG ATA GCA TAC GAT GTT
R G T D L S N L T L I R S D N G W I A Y D V>
6560 6580 6600
TTG TTA ACC AAA GAA GCA GCA AAA GCC TCA CTA CAA TTT GCG TTA AAG AAT CTA CCT AAA GAT GGC
L L T K E A A K A S L Q F A L K N L P K D G>
6620 6640 6660 6680
GAT TTA CCC GTT GTT GCG ATG ATT TAC TCC CAT AGC CAT GCG GAC CAC TTT GGC GGA GCT CGC GGT
D L P V V A M I Y S H S H A D H F G G A R G>
6700 6720 6740
GTT CAA GAG ATG TTC CCT GAT GTC AAA GTC TAC GGC TCA GAT AAC ATC ACT AAA GAA ATT GTC GAT
V Q E M F P D V K V Y G S D N I T K E I V D>

0.92

Fig. 4
4/30

6760 6780 6800
GAG AAC GTA CTT GCC GGT AAC GCC ATG AGC CGC CGC GCA GCT TAT CAA TAC GGC GCA ACA CTG GGC
E N V L A G N A M S R R A A Y Q Y G A T L G>
6820 6840 6860
AAA CAT GAC CAC GGT ATT GTT GAT GCT GCG CTA GGT AAA GGT CTA TCA AAA GGT GAA ATC ACT TAC
K H D H G I V D A A L G K G L S K G E I T Y>
6880 6900 6920 6940
GTC GCC CCA GAC TAC ACC TTA AAC AGT GAA GGC AAA TGG GAA ACG CTG ACG ATT GAT GGT CTA GAG
V A P D Y T L N S E G K W E T L T I D G L E>
6960 6980 7000
ATG GTG TTT ATG GAT GCC TCG GGC ACC GAA GCT GAG TCA GAA ATG ATC ACT TAT ATT CCC TCT AAA
M V F M D A S G T E A E S E M I T Y I P S K>
7020 7040 7060
AAA GCG CTC TGG ACG GCG GAG CTT ACC TAT CAA GGT ATG CAC AAC ATT TAT ACG CTG CGC GGC GCT
K A L W T A E L T Y Q G M H N I Y T L R G A>
7080 7100 7120 7140
AAA GTA CGT GAT GCG CTC AAG TGG TCA AAA GAT ATC AAC GAA ATG ATC AAT GCC TTT GGT CAA GAT
K V R D A L K W S K D I N E M I N A F G Q D>
7160 7180 7200
GTC GAA GTG CTG TTT GCC TCG CAC TCT GCG CCA GTG TGG GGT AAC CAA GCG ATC AAC GAT TTC TTA
V E V L F A S H S A P V W G N Q A I N D F L>
7220 7240 7260
CGC CTA CAG CGT GAT AAC TAC GGC CTA GTG CAC AAT CAA ACC TTG AGA CTT GCC AAC GAT GGT GTC
R L Q R D N Y G L V H N Q T L R L A N D G V>
7280 7300 7320 7340
GGT ATA CAA GAT ATT GGC GAT GCG ATT CAA GAC ACG ATT CCA GAG TCT ATC TAC AAG ACG TGG CAT
G I Q D I G D A I Q D T I P E S I Y K T W H>
7360 7380 7400
ACC AAT GGT TAC CAC GGC ACT TAT AGC CAT AAC GCT AAA GCG GTT TAT AAC AAG TAT CTA GGC TAC
T N G Y H G T Y S H N A K A V Y N K Y L G Y>
7420 7440 7460
TTC GAT ATG AAC CCA GCC AAC CTT AAT CCG CTG CCA ACC AAG CAA GAA TCT GCC AAG TTT GTC GAA
F D M N P A N L N P L P T K Q E S A K F V E>
7480 7500 7520
TAC ATG GGC GGC GCA GAT GCC GCA ATT AAG CGC GCT AAA GAT GAT TAC GCT CAA GGT GAA TAC CGC
Y M G G A D A A I K R A K D D Y A Q G E Y R>
7540 7560 7580 7600
TTT GTT GCA ACG GCA TTA AAT AAG GTG GTG ATG GCC GAG CCA GAA AAT GAC TCC GCT CGT CAA TTG
F V A T A L N K V V M A E P E N D S A R Q L>
7620 7640 7660
CTA GCC GAT ACC TAT GAG CAA CTT GGT TAT CAA GCA GAA GGG GCT GGC TGG AGA AAC ATT TAC TTA
L A D T Y E Q L G Y Q A E G A G W R N I Y L>
7680 7700 7720
ACT GGC GCA CAA GAG CTA CGA GTA GGT ATT CAA GCT GGC GCG CCT AAA ACC GCA TCG GCA GAT GTC
T G A Q E L R V G I Q A G A P K T A S A D V>
7740 7760 7780 7800
ATC AGT GAA ATG GAC ATG CCG ACT CTA TTT GAC TTC CTC GCG GTG AAG ATT GAT AGT CAA CAG GCG
I S E M D M P T L F D F L A V K I D S Q Q A>
7820 7840 7860
GCT AAG CAC GGC TTA GTT AAG ATG AAT GTT ATC ACC CCT GAT ACT AAA GAT ATT CTC TAT ATT GAG
A K H G L V K M N V I T P D T K D I L Y I E>
7880 7900 7920
CTA AGC AAC GGT AAC TTA AGC AAC GCA GTG GTC GAC AAA GAG CAA GCA GCT GAC GCA AAC CTT ATG
L S N G N L S N A V V D K E Q A A D A N L M>

Fig. 4
5/30

7940 7960 7980 8000
GTT AAT AAA GCT GAC GTT AAC CGC ATC TTA CTT GGC CAA GTA ACC CTA AAA GCG TTA TTA GCC AGC
V N K A D V N R I L L G Q V T L K A L L A S>
8020 8040 8060
GGC GAT GCC AAG CTC ACT GGT GAT AAA ACG GCA TTT ACT AAA ATA GCC GAT AGC ATG GTC GAG TTT
G D A K L T G D K T A F S K I A D S M V E F>
8080 8100 8120 8140
ACA CCT GAC TTC GAA ATC GTA CCA ACG CCT GTT AAA TGAGGCA TTAATCTCAA CAAGTGCAAG CTAGACATAA
T P D F E I V P T P V K>
8160 8180 8200
AAATGGGGCG ATTAGACGCC CCATTTTITA TGCAATTTTG AACTA GCT AGT CTT AGC TGA AGC TCG AAC AAC
<S T K A S A R V V
8220 8240 8260
AGC TTT AAA ATT CAC TTC TTC TGC TGC AAT ACT TAT TTG CTG ACA CTG ACC AAT ACT CAG TGC AAA
<A K F N V E E A A I S I Q Q C Q G I S L A F
8280 8300 8320 8340
ACG ATA ACT ATC ATC AAG ATG GCC CAG TAA ACA ATG CCA ATT ATC AGC AGC GTT CAT TTG CTG TTC
<R Y S D D L H G L L C H W N D A A N M Q Q E
8360 8380 8400
TTT AGC CTC AAT CAA ACC TAA ACC AGA CTT TTG TGG CTC AGC CTT AGG CTT ATT AGA ACT CGA CTC
<K A E I L G L G S K Q P E A N P K N S S S E
8420 8440 8460
TAG TAA AGC AAG ACC AAT ATC TTG TTT TAA CAA AAC CTG TCG CTG ATT AAG TTG ATG CTC AAC CTT
<L L A L G I D Q K L L V Q R Q N L Q H E V K
8480 8500 8520 8540
GTG ATC CGC AAT AGC ATC GGA AAT ATC AAC ACA ATG GCT CAA GCT TTT AGG TGC ATT AAC TCC AAG
<H D A I A D S I D V C H S L S K P A N V G L
8560 8580 8600
AAA AGT TTC GCT CAG TGC AGA GAA GTC AAA CGC AAA AGA TTT TAG CGA TAA TGC CAG CCC AAG TCC
<F T E S L A S F D F A F S K L S L A L G L G
8620 8640 8660
TTT CGC TTT AAT GTA AGA CTC CTT GAG CGC CCA CAA ATC AAA AAA GCG GTC TCG CTG CAA GGC CTC
<K A K I Y S E K L A W L D F F R D R Q L A E
8680 8700 8720 8740
TGG TAA CGC TAA CAA GGC TCG CTT TTC TGA TTC AGA GAA ATA ATG ACT AAG AAT AGA GTG GAT ATT
<P L A L L A R K E S E S F Y H S L I S H I N
8760 8780 8800
GGT GCT GTT ACG GCA ACG CTC AAT GTC GAC GCC AAA CTC AAT ACT AGC AGA GTC AGT TTC CTC CTT
<T S N R C R E I D V G F E I S A S D T E E K
8820 8840 8860
GCT TGC CTG ACT GGC GCC TTT ATT ATC AGC AGT GCA AAT GCC TAC TAA TAG CCA ATC TCC ACT ATG
<S A Q S A G K N D A T C I G V L L W D G S H
8880 8900 8920
ACT CAC ATT AAA GTG GAC CCC GGT TTG AGC AAA TTG CGC ATC ACT CAA TCT AGG CTT ACC TTT GTC
<S V N F H V G T Q A F Q A D S L R P K G K D
8940 8960 8980 9000
GCC ATA TTC AAA GCG CCA TTC ATT GGG GCG TAT TTC ACT ATG TTG TGA CAA TAA AGC GCG CAA ATA
<G Y E F R W E N P R I E S H Q S L L A R L Y
9020 9040 9060 9080
GCC TCT TAC CAT TAAA CCTTGAGTTT TAGCTTCTTG TTTAATGTAG CGATTAACCT TAATTAACCT ATCTTCAGGC
<G R V M
9100 9120 9140 9160
AGCCATGACT TAACCAACTC TGAGTCTGG TTATCGCACT CTTGTATTGT TAACGGACAG AAGTATAAGG AAATCAATCG

0-43
Fig. 4
6/30

9180 9200 9220 9240
AGAAGTTAGC AATTTTTCAG GACACTCTTT AAAGCAACAA ACATAACCCC TATTTTACC AATTTAAGAT CAAAACATAA
9260 9280 9300 9320
GCCAAACTA ATTGAGAATA GTGTCAAAC AGCTTTAAAG GAAAAAATA TAAAAAGAAC ATTATACTTG TATAAATTAT
9340 9360 9380 9400
TTTACACACC AAAGCCATGA TCTTCACAAA ATTAGCTCCC TCTCCCTAAA ACAAGATTGA ATAAAAAAT AACCTTAAC
9420 9440 9460 9480
TTTCATATAG ATAAAAACAAA CCAATGGGAT AAAGTATATT GAATTCATTT TTAAGGAAAA ATTCAAATTG AATTCAAGCT
9500 9520 9540 9560
CTTCAGTAAA AGCATATTTT GCCGTTAGTG TGAAAAAATA CAAATTTAAA AACCAACATA GAACAAATAA GCAGACAATA
9580 9600 9620 9640
AAACCAAGGC GCAACACAAA CAACGCGCTT ACAATTTTCA CAAAAAGCA ACAAGAGTAA CGTTTAGTAT TTGGATATGG
9660 9680 9700
TTATTGTAAT TGAGAATTTT ATAACAATTA TATTAAGGGA ATG AGT ATG TTT TTA AAT TCA AAA CTT TCG CGC
M S M F L N S K L S R>
9720 9740 9760
TCA GTC AAA CTT GCC ATA TCC GCA GGC TTA ACA GCC TCG CTA GCT ATG CCT GTT TTT GCA GAA GAA
S V K L A I S A G L T A S L A M P V F A E E>
9780 9800 9820 9840
ACT GCT GCT GAA GAA CAA ATA GAA AGA GTC GCA GTG ACC GGA TCG CGA ATC GCT AAA GCA GAG CTA
T A A E E Q I E R V A V T G S R I A K A E L>
9860 9880 9900
ACT CAA CCA GCT CCA GTC GTC AGC CTT TCA GCC GAA GAA CTG ACA AAA TTT GGT AAT CAA GAT TTA
T Q P A P V V S L S A E E L T K F G N Q D L>
9920 9940 9960
GGT AGC GTA CTA GCA GAA TTA CCT GCT ATT GGT GCA ACC AAC ACT ATT ATT GGT AAT AAC AAT AGC
G S V L A E L P A I G A T N T I I G N N N S>
9980 10000 10020 10040
AAC TCA AGC GCA GGT GTT AGC TCA GCA GAC TTG CGT CGT CTA GGT GCT AAC AGA ACC TTA GTA TTA
N S S A G V S S A D L R R L G A N R T L V L>
10060 10080 10100
GTC AAC GGT AAG CGC TAC GTT GCC GGC CAA CCG GGC TCA GCT GAG GTA GAT TTG TCA ACT ATA CCA
V N G K R Y V A G Q P G S A E V D L S T I P>
10120 10140 10160
ACT AGC ATG ATC TCG CGA GTT GAG ATT GTA ACC GGC GGT GCT TCA GCA ATT TAT GGT TCG GAC GCT
T S M I S R V E I V T G G A S A I Y G S D A>
10180 10200 10220 10240
GTA TCA GGT GTT ATC AAC GTT ATC CTT AAA GAA GAC TTT GAA GGC TTT GAG TTT AAC GCA CGT ACT
V S G V I N V I L K E D F E G F E F N A R T>
10260 10280 10300
AGC GGT TCT ACT GAA AGT GTA GGC ACT CAA GAG CAC TCT TTT GAC ATT TTG GGT GGT GCA AAC GTT
S G S T E S V G T Q E H S F D I L G G A N V>
10320 10340 10360
GCA GAT GGA CGT GGT AAT GTA ACC TTC TAC GCA GGT TAT GAA CGT ACA AAA GAA GTC ATG GCT ACC
A D G R G N V T F Y A G Y E R T K E V M A T>
10380 10400 10420
GAC ATT CGC CAA TTC GAT GCT TGG GGA ACA ATT AAA AAC GAA GCC GAT GGT GGT GAA GAT GAT GGT
D I R Q F D A W G T I K N E A D G G E D D G>
10440 10460 10480 10500
ATT CCA GAC ACA CTA CGT GTA CCA CGA GTT TAT TCT GAA ATG ATT AAT GCT ACC GGT GTT ATC AAT
I P D R L R V P R V Y S E M I N A T G V I N>

0-642

Fig. 4
7/30

10520 10540 10560
GCA TTT GGT GGT GGA ATT GGT CGC TCA ACC TTT GAC AGT AAC GGC AAT CCT ATT GCA CAA CAA GAA
A F G G G I G R S T F D S N G N P I A Q Q E>
10580 10600 10620
CGT GAT GGG ACT AAC AGC TTT GCA TTT GGT TCA TTC CCT AAT GGC TGT GAC ACA TGT TTC AAC ACT
R D G T N S F A F G S F P N G C D T C F N T>
10640 10660 10680 10700
GAA GCA TAC GAA AAC TAT ATT CCA GGG GTA GAA AGA ATA AAC GTT GGC TCA TCA TTC AAC TTT GAT
E A Y E N Y I P G V E R I N V G S S F N F D>
10720 10740 10760
TTT ACC GAT AAC ATT CAA TTT TAC ACT GAC TTC AGA TAT GTA AAG TCA GAT ATT CAG CAA CAA TTT
P T D N I Q F Y T D F R Y V K S D I Q Q Q F>
10780 10800 10820
CAG CCT TCA TTC CGT TTT GGT AAC ATT AAT ATC AAT GTT GAA GAT AAC GCC TTT TTG AAT GAC GAC
Q P S F R F G N I N I N V E D N A F L N D D>
10840 10860 10880 10900
TTG CGT CAG CAA ATG CTC GAT GCG GGT CAA ACC AAT GCT AGT TTT GCC AAG TTT TTT GAT GAA TTA
L R Q Q M L D A G Q T N A S F A K F F D E L>
10920 10940 10960
GGA AAT CGC TCA GCA GAA AAT AAA CGC GAA CTT TTC CGT TAC GTA GGT GGC TTT AAA GGT GGC TTT
G N R S A E N K R E L F R Y V G G F K G G F>
10980 11000 11020
GAT ATT AGC GAA ACC ATA TTT GAT TAC GAC CTT TAC TAT GTT TAT GGC GAG ACT AAT AAC CGT CGT
D I S E T I F D Y D L Y Y V Y G E T N N R R>
11040 11060 11080
AAA ACC CTT AAT GAC CTA ATT CCT GAT AAC TTT GTC GCA GCT GTC GAC TCT GTT ATT GAT CCT GAT
K T L N D L I P D N F V A A V D S V I D P D>
11100 11120 11140 11160
ACT GGC TTA GCA GCG TGT CGC TCA CAA GTA GCA AGC GCT CAA GGC GAT GAC TAT ACA GAT CCC GCG
T G L A A C R S Q V A S A Q G D D Y T D P A>
11180 11200 11220
TCT GTA AAT GGT AGC GAC TGT GTT GCT TAT AAC CCA TTT GGC ATG GGT CAA GCT TCA GCA GAA GCC
S V N G S D C V A Y N P F G M G Q A S A E A>
11240 11260 11280
CGC GAC TGG GTT TCT GCT GAT GTG ACT CGT GAA GAC AAA ATA ACT CAA CAA GTG ATT GGT GGT ACT
R D W V S A D V T R E D K I T Q Q V I G G T>
11300 11320 11340 11360
CTC GGT ACC GAT TCT GAA GAA CTA TTT GAG CTT CAA GGT GGT GCA ATC GCT ATG GTT GTT GGT TTT
L G T D S E E L F E L Q G G A I A M V V G F>
11380 11400 11420
GAA TAC CGT GAA GAA ACG TCT GGT TCA ACA ACC GAT GAA TTT ACT AAA GCA GGT TTC TTG ACA AGC
E Y R E E T S G S T T D E F T K A G F L T S>
11440 11460 11480
GCT GCA ACG CCA GAT TCT TAT GGC GAA TAC GAC GTG ACT GAG TAT TTT GTT GAG GTG AAC ATC CCA
A A T P D S Y G E Y D V T E Y F V E V N I P>
11500 11520 11540 11560
GTA CTA AAA GAA TTA CCT TTT GCA CAT GAG TTG AGC TTT GAC GGT GCA TAC CGT AAT GCT GAT TAC
V L K E L P F A H E L S F D G A Y R N A D Y>
11580 11600 11620
TCA CAT GCC GGT AAG ACT GAA GCA TGG AAA GCT GGT ATG TTC TAC TCA CCA TTA GAG CAA CTT GCA
S H A G K T E A W K A G M F Y S P L E Q L A>
11640 11660 11680
TTA CGT GGT ACG GTA GGT GAA GCA GTA CGA GCA CCA AAC ATT GCA GAA GCC TTT AGT CCA CGC TCT

Fig. 4
8/30

L R G T V G E A V R A P N I A E A F S P R S>
11700 11720 11740
CCT GGT TTT GGC CGC GTT TCA GAT CCA TGT GAT GCA GAT AAC ATT AAT GAC GAT CCG GAT CGC GTG
P G F G R V S D P C D A D N I N D D P D R V>
11760 11780 11800 11820
TCA AAC TGT GCA GCA TTG GGG ATC CCT CCA GGA TTC CAA GCT AAT GAT AAC GTC AGT GTA GAT ACC
S N C A A L G I P P G F Q A N D N V S V D T>
11840 11860 11880
TTA TCT GGT GGT AAC CCA GAT CTA AAA CCT GAA ACA TCA ACA TCC TTT ACA GGT GGT CTT GTT TGG
L S G G N P D L K P E T S T S F T G G L V W>
11900 11920 11940
ACA CCA ACG TTT GCT GAC AAT CTA TCA TTC ACT GTC GAT TAT TAT GAT ATT CAA ATT GAG GAT GCT
T P T F A D N L S F T V D Y Y D I Q I E D A>
11960 11980 12000 12020
ATT TTG TCA GTA GCC ACC CAG ACT GTG GCT GAT AAC TGT GTT GAC TCA ACT GGC GGA CCT GAC ACC
I L S V A T Q T V A D N C V D S T G G P D T>
12040 12060 12080
GAC TTC TGT AGT CAA GTT GAT CGT AAT CCA ACG ACC TAT GAT ATT GAA CTT GTT CGC TCT GGT TAT
D F C S Q V D R N P T T Y D I E L V R S G Y>
12100 12120 12140
CTA AAT GCC GCG GCA TTG AAT ACC AAA GGT ATT GAA TTT CAA GCT GCA TAC TCA TTA GAT CTA GAG
L N A A A L N T K G I E F Q A A Y S L D L E>
12160 12180 12200 12220
TCT TTC AAC GCG CCT GGT GAA CTA CGC TTC AAC CTA TTG GGG AAC CAA TTA CTT GAA CTA GAA CGT
S F N A P G E L R F N L L G N Q L L E L E R>
12240 12260 12280
CTT GAA TTC CAA AAT CGT CCT GAT GAG ATT AAT GAT GAA AAA GGC GAA GTA GGT GAT CCA GAG CTG
L E F Q N R P D E I N D E K G E V G D P E L>
12300 12320 12340
CAG TTC CGC CTA GGC ATC GAT TAC CGT CTA GAT GAT CTA AGT GTT AGC TGG AAC ACG CGT TAT ATT
Q F R L G I D Y R L D D L S V S W N T R Y I>
12360 12380 12400
GAT AGC GTA GTA ACT TAT GAT GTC TCT GAA AAT GGT GGC TCT CCT GAA GAT TTA TAT CCA GGC CAC
D S V V T Y D V S E N G G S P E D L Y P G H>
12420 12440 12460 12480
ATA GGC TCA ATG ACA ACT CAT GAC TTG AGC GCT ACA TAC TAC ATC AAT GAG AAC TTC ATG ATT AAC
I G S M T T H D L S A T Y Y I N E N F M I N>
12500 12520 12540
GGT GGT GTA CGT AAC CTA TTT GAC GCA CTT CCA CCT GGA TAC ACT AAC GAT GCG CTA TAT GAT CTA
G G V R N L F D A L P P G Y T N D A L Y D L>
12560 12580 12600 12620
GTT GGT CGC CGT GCA TTC CTA GGT ATT AAG GTA ATG ATG TAATTAATTA TTACGCCTCT AACTAATAAA
V G R R A F L G I K V M M>
12640 12660 12680 12700
AATGCAATCT CTTCGTAGAG ATTGCATTTT TTTATGAAAT CCAATCTTAA ACTGGTTCTC CGAGCATCTT ACGCCTTAAA
12720 12740 12760 12780
AACCCCGCCC CTCAATGTAA CGCCAAAGTT AATTGCTTAC ACGCACTTAC ACAACGAAC AATTTCATTA ACACGAGACA
12800 12820 12840 12860
CAGCTCAGCG TTTTATTTT ACCCTTGATT TTAACATATA AAATTGCGTT TTAGCGCACA AGTGTCTCC CAAGCTGGTC
12880 12900 12920 12940
GTATCTGTAA TTATTCAGTC CCAGGTGATT GTATTGACCC ATAAGCTCAG GTAGTCTGCT CTGCCATTAG CTAACAATA
12960 12980 13000 13020

look for Fur
Fig. 4
9/30

TTGACAAAAT GGCATAAAA TGTGGCTTAG CGCTAAGTTC ACCGTAAGTT TTATCGGCAT TAAGTCCCAA CAGATTATTA
13040 13060 13080
ACGGAAACCC GCTAAACTG ATC GCA AAA ATA AAT AGT GAA CAC TTG GAT GAA GCT ACT ATT ACT TCG AAT
M A K I N S E H L D E A T I T S N>
13100 13120 13140
AAG TGT ACG CAA ACA GAG ACT GAG GCT CGG CAT AGA AAT GCC ACT ACA ACA CCT GAG ATG CGC CGA
K C T Q T E T E A R H R N A T T T P E M R R>
13160 13180 13200 13220
TTC ATA CAA GAG TCG GAT CTC AGT GTT AGC CAA CTG TCT AAA ATA TTA AAT ATC AGT GAA GCT ACC
F I Q E S D L S V S Q L S K I L N I S E A T>
13240 13260 13280
GTA CGT AAG TGG CGC AAG CGT GAC TCT GTC GAA AAC TGT CCT AAT ACC CCG CAC CAT CTC AAT ACC
V R K W R K R D S V E N C P N T P H H L N T>
13300 13320 13340
ACG CTA ACC CCT TTG CAA GAA TAT GTG GTT GTG GGC CTG CGT TAT CAA TTG AAA ATG CCA TTA GAC
T L T P L Q E Y V V G L R Y Q L K M P L D>
13360 13380 13400 13420
AGA TTG CTC AAA GCA ACC CAA GAG TTT ATC AAT CCA AAC GTG TCG CGC TCA GGT TTA GCA AGA TGT
R L L K A T Q E F I N P N V S R S G L A R C>
13440 13460 13480
TTG AAG CGT TAT GGC GTT TCA CGG GTG AGT GAT ATC CAA AGC CCA CAC GTA CCA ATG CGC TAC TTT
L K R Y G V S R V S D I Q S P H V P M R Y F>
13500 13520 13540
AAT CAA ATT CCA GTC ACT CAA GGC AGC GAT GTG CAA ACC TAC ACC CTG CAC TAT GAA ACG CTG GCA
N Q I P V T Q G S D V Q T Y T L H Y E T L A>
13560 13580 13600
AAA ACC TTA GCC TTA CCT AGT ACC GAT GGT GAC AAT GTG GTG CAA GTG GTG TCT CTC ACC ATT CCA
K T L A L P S T D G D N V V Q V V S L T I P>
13620 13640 13660 13680
CCA AAG TTA ACC GAA GAA GCA CCC AGT TCA ATT TTG CTC GGC ATT GAT CCT CAT AGC GAC TGG ATC
P K L T E E A P S S I L L G I D P H S D W I>
13700 13720 13740
TAT CTC GAC ATA TAC CAA GAT GGC AAT ACA CAA GCC ACG AAT AGA TAT ATG GCT TAT GTG CTA AAA
Y L D I Y Q D G N T Q A T N R Y M A Y V L K>
13760 13780 13800
CAC GGG CCA TTC CAT TTA CGA AAG TTA CTC GTG CGT AAC TAT CAC ACC TTT TTA CAG CGC TTT CCT
H G P F H L R K L V R N Y H T F L Q R F P>
13820 13840 13860 13880
GGA GCG ACG CAA AAT CGC CGC CCC TCT AAA GAT ATG CCT GAA ACA ATC AAC AAG ACG CCT GAA ACA
G A T Q N R R P S K D M P E T I N K T P E T>
13900 13920 13940
CAG GCA CCC AGT GGA GAC TCA TA ATC AGC CAG ACC TCT AAA CCT ACA AAC TCA GCA ACT GAG CAA
Q A P S G D S> M S Q T S K P T N S A T E Q>
13960 13980 14000
GCA CAA GAC TCA CAA GCT GAC TCT CGT TTA AAT AAA CGA CTA AAA CAT ATG CCA ATT GCT ATT GTT
A Q D S Q A D S R L N K R L K D M P I A I V>
14020 14040 14060
GGC ATG GCG AGT ATT TTT GCA AAC TCT CGC TAT TTG AAT AAG TTT TGG GAC TTA ATC AGC GAA AAA
G M A S I F A N S R Y L N K F W D L I S E K>
14080 14100 14120 14140
ATT GAT GCG ATT ACT GAA TTA CCA TCA ACT CAC TGG CAG CCT GAA GAA TAT TAC GAC GCA GAT AAA
I D A I T E L P S T H W Q P E E Y Y D A D K>

o-f 4

o-f 6

Fig. 4
10/30

14160 14180 14200
ACC GCA GCA GAC AAA AGC TAC TGT AAA CGT GGT GGC TTT TTG CCA GAT GTA GAC TTC AAC CCA ATG
T A A D K S Y C K R G G F L P D V D F N P M>
14220 14240 14260
GAG TTT GGC CTG CCG CCA AAC ATT TTG GAA CTG ACC GAT TCA TCG CAA CTA TTA TCA CTC ATC GTT
E F G L P P N I L E L T D S S Q L L S L I V>
14280 14300 14320 14340
GCT AAA GAA GTG TTG GCT GAT GCT AAC TTA CCT GAG AAT TAC GAC CGC GAT AAA ATT GGT ATC ACC
A K E V L A D A N L P E N Y D R D K I G I T>
14360 14380 14400
TTA GGT GTC GGC GGT GGT CAA AAA ATT AGC CAC AGC CTA ACA GCG CGT CTG CAA TAC CCA GTA TTG
L G V G G G Q K I S H S L T A R L Q Y P V L>
14420 14440 14460
AAG AAA GTA TTC GCC AAT AGC GGC ATT AGT GAC ACC GAC AGC GAA ATG CTT ATC AAG AAA TTC CAA
K K V F A N S G I S D T D S E M L I K K F Q>
14480 14500 14520 14540
GAC CAA TAT GTA CAC TGG GAA AAC TCG TTC CCA GGT TCA CTT GGT AAC GTT ATT GCG GGC CGT
D Q Y V H W E N S F P G S L G N V I A G R>
14560 14580 14600
ATC GCC AAC CGC TTC GAT TTT GGC GGC ATG AAC TGT GTG GTT GAT GCT GCC TGT GCT GGA TCA CTT
I A N R F D F G G M N C V V D A A C A G S L>
14620 14640 14660
GCT GCT ATG CGT ATG GCG CTA ACA GAG CTA ACT GAA GGT CGC TCT GAA ATG ATG ATC ACC GGT GGT
A A M R M A L T E L T E G R S E M M I T G G>
14680 14700 14720
GTG TGT ACT GAT AAC TCA CCC TCT ATG TAT ATG AGC TTT TCA AAA ACG CCC GCC TTT ACC ACT AAC
V C T D N S P S M Y M S F S K T P A F T T N>
14740 14760 14780 14800
GAA ACC ATT CAG CCA TTT GAT ATC GAC TCA AAA GGC ATG ATG ATT GGT GAA GGT ATT GGC ATG GTG
E T I Q P F D I D S K G M M I G E G I G M V>
14820 14840 14860
GCG CTA AAG CGT CTT GAA GAT GCA GAG CGC GAT GGC GAC CGC ATT TAC TCT GTA ATT AAA GGT GTG
A L K R L E D A E R D G D R I Y S V I K G V>
14880 14900 14920
GGT GCA TCA TCT GAC GGT AAG TTT AAA TCA ATC TAT GCC CCT CGC CCA TCA GGC CAA GCT AAA GCA
G A S S D G K F K S I Y A P R P S G Q A K A>
14940 14960 14980 15000
CTT AAC CGT GCC TAT GAT GAC GCA GGT TTT GCG CCG CAT ACC TTA GGT CTA ATT GAA GCT CAC GGA
L N R A Y D D A G F A P H T L G L I E A H G>
15020 15040 15060
ACA GGT ACT GCA GCA GGT GAC GCG GCA GAG TTT GCC GGC CTT TGC TCA GTA TTT GCT GAA GGC AAC
T G T A A G D A A E F A G L C S V F A E G N>
15080 15100 15120
GAT ACC AAG CAA CAC ATT GCG CTA GGT TCA GTT AAA TCA CAA ATT GGT CAT ACT AAA TCA ACT GCA
D T K Q H I A L G S V K S Q I G H T K S T A>
15140 15160 15180 15200
GGT ACA GCA GGT TTA ATT AAA GCT GCT CTT GCT TTG CAT CAC AAG GTA CTG CCG CCG ACC ATT AAC
G T A G L I K A A L A L H H K V L P P T I N>
15220 15240 15260
GTT AGT CAG CCA AGC CCT AAA CTT GAT ATC GAA AAC TCA CCG TTT TAT CTA AAC ACT GAG ACT CGT
V S Q P S P K L D I E N S P F Y L N T E T R>
15280 15300 15320
CCA TGG TTA CCA CGT GTT GAT GGT ACG CCG CGC CGC GCG GGT ATT AGC TCA TTT GGT TTT GGT GGC
P W L P R V D G T P R R A G I S S F G F G G>

Fig. 4
11/30

15340 15360 15380
ACT AAC TTC CAT TTT GTA CTA GAA GAG TAC AAC CAA GAA CAC AGC CGT ACT GAT AGC GAA AAA GCT
T N F H F V L E E Y N Q E H S R T D S E K A>
15400 15420 15440 15460
AAG TAT CGT CAA CGC CAA GTG GCG CAA AGC TTC CTT GTT AGC GCA AGC GAT AAA GCA TCG CTA ATT
K Y R Q R Q V A Q S F L V S A S D K A S L I>
15480 15500 15520
AAC GAG TTA AAC GTA CTA GCA GCA TCT GCA AGC CAA GCT GAG TTT ATC CTC AAA GAT GCA GCA GCA
N E L N V L A A S A S Q A E F I L K D A A A>
15540 15560 15580
AAC TAT GGC GTA CGT GAG CTT GAT AAA AAT GCA CCA CGG ATC GGT TTA GTT GCA AAC ACA GCT GAA
N Y G V R E L D K N A P R I G L V A N T A E>
15600 15620 15640 15660
GAG TTA GCA GGC CTA ATT AAG CAA GCA CTT GCC AAA CTA GCA GCT AGC GAT GAT AAC GCA TGG CAG
E L A G L I K Q A L A K L A A S D D N A W Q>
15680 15700 15720
CTA CCT GGT GGC ACT AGC TAC CGC GCC GCT GCA GTA GAA GGT AAA GTT GCC GCA CTG TTT GCT GGC
L P G G T S Y R A A A V E G K V A A L F A G>
15740 15760 15780
CAA GGT TCA CAA TAT CTC AAT ATG GGC CGT GAC CTT ACT TGT TAT TAC CCA GAG ATG CGT CAG CAA
Q G S Q Y L N M G R D L T C Y Y P E M R Q Q>
15800 15820 15840 15860
TTT GTA ACT GCA GAT AAA GTA TTT GCC GCA AAT GAT AAA ACG CCG TTA TCG CAA ACT CTG TAT CCA
F V T A D K V F A A N D K T P L S Q T L Y P>
15880 15900 15920
AAG CCT GTA TTT AAT AAA GAT GAA TTA AAG GCT CAA GAA GCC ATT TTG ACC AAT ACC GCC AAT GCC
K P V F N K D E L K A Q E A I L T N T A N A>
15940 15960 15980
CAA AGC GCA ATT GGT GCG ATT TCA ATG GGT CAA TAC GAT TTG TTT ACT GCG GCT GGC TTT AAT GCC
Q S A I G A I S M G Q Y D L F T A A G F N A>
16000 16020 16040
GAC ATG GTT GCA GGC CAT AGC TTT GGT GAG CTA AGT GCA CTG TGT GCT GCA GGT GTT ATT TCA GCT
D M V A G H S F G E L S A L C A A G V I S A>
16060 16080 16100 16120
GAT GAC TAC TAC AAG CTG GCT TTT GCT CGT GGT GAG GCT ATG GCA ACA AAA GCA CCG GCT AAA GAC
D D Y Y K L A F A R G E A M A T K A P A K D>
16140 16160 16180
GGC GTT GAA GCA GAT GCA GGA GCA ATG TTT GCA ATC ATA ACC AAG AGT GCT GCA GAC CTT GAA ACC
G V E A D A G A M F A I I T K S A A D L E T>
16200 16220 16240
GTT GAA GCC ACC ATC GCT AAA TTT GAT GGG GTG AAA GTC GCT AAC TAT AAC GCG CCA ACG CAA TCA
V E A T I A K F D G V K V A N Y N A P T Q S>
16260 16280 16300 16320
GTA ATT GCA GGC CCA ACA GCA ACT ACC GCT GAT GCG GCT AAA GCG CTA ACT GAG CTT GGT TAC AAA
V I A G P T A T T A D A A K A L T E L G Y K>
16340 16360 16380
GCG ATT AAC CTG CCA GTA TCA GGT GCA TTC CAC ACT GAA CTT GTT GGT CAC GCT CAA GCG CCA TTT
A I N L P V S G A F H T E L V G H A Q A P F>
16400 16420 16440
GCT AAA GCG ATT GAC GCA GCC AAA TTT ACT AAA ACA AGC CGA GCA CTT TAC TCA AAT GCA ACT GGC
A K A I D A A K F T K T S R A L Y S N A T G>
16460 16480 16500 16520
GGA CTT TAT GAA AGC ACT GCT GCA AAG ATT AAA GCC TCG TTT AAG AAA CAT ATG CTT CAA TCA GTG

Fig. 4
12/30

G L Y E S T A A K I K A S F K K H M L Q S V>
16540 16560 16580
CGC TTT ACT AGC CAG CTA GAA GCC ATG TAC AAC GAC GGC GCC CGT GTA TTT GTT GAA TTT GGT CCA
R F T S Q L E A M Y N D G A R V F V E F G P>
16600 16620 16640
AAG AAC ATC TTA CAA AAA TTA GTT CAA GGC ACG CTT GTC AAC ACT GAA AAT GAA GTT TGC ACT ATC
K N I L Q K L V Q G T L V N T E N E V C T I>
16660 16680 16700
TCT ATC AAC CCT AAT CCT AAA GTT GAT AGT GAT CTG CAG CTT AAG CAA GCA GCA ATG CAG CTA GCG
S I N P N P K V D S D L Q L K Q A A M Q L A>
16720 16740 16760 16780
GTT ACT GGT GTG GTA CTC AGT GAA ATT GAC CCA TAC CAA GCC GAT ATT GCC GCA CCA GCG AAA AAG
V T G V V L S E I D P Y Q A D I A A P A K K>
16800 16820 16840
TCG CCA ATG AGC ATT TCG CTT AAT GCT GCT AAC CAT ATC AGC AAA GCA ACT CGC GCT AAG ATG GCC
S P M S I S L N A A N H I S K A T R A K M A>
16860 16880 16900
AAG TCT TTA GAG ACA GGT ATC GTC ACC TCG CAA ATA GAA CAT GTT ATT GAA GAA AAA ATC GTT GAA
K S L E T G I V T S Q I E H V I E E K I V E>
16920 16940 16960 16980
GTT GAG AAA CTG GTT GAA GTC GAA AAG ATC GTC GAA AAA GTG GTT GAA GTA GAG AAA GTT GTT GAG
V E K L V E V E K I V E K V V E V E K V V E>
17000 17020 17040
GTT GAA GCT CCT GTT AAT TCA GTG CAA GCC AAT GCA ATT CAA ACC CGT TCA GTT GTC GCT CCA GTA
V E A P V N S V Q A N A I Q T R S V V A P V>
17060 17080 17100
ATA GAG AAC CAA GTC GTG TCT AAA AAC AGT AAG CCA GCA GTC CAG AGC ATT AGT GGT GAT GCA CTC
I E N Q V V S K N S K P A V Q S I S G D A L>
17120 17140 17160 17180
AGC AAC TTT TTT GCT GCA CAG CAG CAA ACC GCA CAG TTG CAT CAG CAG TTC TTA GCT ATT CCG CAG
S N F F A A Q Q Q T A Q L H Q Q F L A I P Q>
17200 17220 17240
CAA TAT GGT GAG ACG TTC ACT ACG CTG ATG ACC GAG CAA GCT AAA CTG GCA AGT TCT GGT GTT GCA
Q Y G E T F T T L M T E Q A K L A S S G V A>
17260 17280 17300
ATT CCA GAG AGT CTG CAA CGC TCA ATG GAG CAA TTC CAC CAA CTA CAA GCG CAA ACA CTA CAA AGC
I P E S L Q R S M E Q F H Q L Q A Q T L Q S>
17320 17340 17360
CAC ACC CAG TTC CTT GAG ATG CAA GCG GGT AGC AAC ATT GCA GCG TTA AAC CTA CTC AAT AGC AGC
H T Q F L E M Q A G S N I A A L N L L N S S>
17380 17400 17420 17440
CAA GCA ACT TAC GCT CCA GCC ATT CAC AAT GAA GCG ATT CAA AGC CAA GTG GTT CAA AGC CAA ACT
Q A T Y A P A I H N E A I Q S Q V V Q S Q T>
17460 17480 17500
GCA GTC CAG CCA GTA ATT TCA ACA CAA GTT AAC CAT GTG TCA GAG CAG CCA ACT CAA GCT CCA GCT
A V Q P V I S T Q V N H V S E Q P T Q A P A>
17520 17540 17560
CCA AAA GCG CAG CCA GCA CCT GTG ACA ACT GCA GTT CAA ACT GCT CCG GCA CAA GTT GTT CGT CAA
P K A Q P A P V T T A V Q T A P A Q V V R Q>
17580 17600 17620 17640
GCC GCA CCA GTT CAA GCC GCT ATT CAA CCG ATT AAT ACA AGT GTT GCG ACT ACA ACG CCT TCA GCC
A A P V Q A A I E P I N T S V A T T T P S A>
17660 17680 17700

Fig. 4
13/30

TTC AGC GCC GAA ACA GCC CTG AGC GCA ACA AAA GTC CAA GCC ACT ATG CTT GAA GTG GTT GCT GAG
F S A E T A L S A T K V Q A T M L E V V A E>

17720 17740 17760
AAA ACC GGT TAC CCA ACT GAA ATG CTA GAG CTT GAA ATG GAT ATG GAA GCC GAT TTA GGC ATC GAT
K T G Y P T E M L E L E M D M E A D L G I D>

17780 17800 17820 17840
TCT ATC AAG CGT GTA GAA ATT CTT GGC ACA GTA CAA GAT GAG CTA CCG GGT CTA CCT GAG CTT AGC
S I K R V E I L G T V Q D E L P G L P E L S>

17860 17880 17900
CCT GAA GAT CTA GCT GAG TGT CGA ACG CTA GGC GAA ATC GTT GAC TAT ATG GGC AGT AAA CTG CCG
P E D L A E C R T L G E I V D Y M G S K L P>

17920 17940 17960
GCT GAA GGC TCT ATG AAT TCT CAG CTG TCT ACA GGT TCC GCA GCT GCG ACT CCT GCA GCG AAT GGT
A E G S M N S Q L S T G S A A A T P A A N G>

17980 18000 18020
CTT TCT GCG GAG AAA GTT CAA GCG ACT ATG ATG TCT GTG GTT GCC GAA AAG ACT GGC TAC CCA ACT
L S A E K V Q A T M M S V V A E K T G Y P T>

18040 18060 18080 18100
GAA ATG CTA GAG CTT GAA ATG GAT ATG GAA GCC GAT TTA GGC ATA GAT TCT ATC AAG CGC GTT GAA
E M L E L E M D M E A D L G I D S I K R V E>

18120 18140 18160
ATT CTT GGC ACA GTA CAA GAT GAG CTA CCG GGT CTA CCT GAG CTT AGC CCT GAA GAT CTA GCT GAG
I L G T V Q D E L P G L P E L S P E D L A E>

18180 18200 18220
TGT CGT ACT CTA GGC GAA ATC GTT GAC TAT ATG AAC TCT AAA CTC GCT GAC GGC TCT AAG CTG CCG
C R T L G E I V D Y M N S K L A D G S K L P>

18240 18260 18280 18300
GCT GAA GGC TCT ATG AAT TCT CAG CTG TCT ACA AGT GCC GCA GCT GCG ACT CCT GCA GCG AAT GGT
A E G S M N S Q L S T S A A A A T P A A N G>

18320 18340 18360
CTC TCT GCG GAG AAA GTT CAA GCG ACT ATG ATG TCT GTG GTT GCC GAA AAG ACT GGC TAC CCA ACT
L S A E K V Q A T M M S V V A E K T G Y P T>

18380 18400 18420
GAA ATG CTA GAA CTT GAA ATG GAT ATG GAA GCT GAC CTT GGC ATC GAT TCA ATC AAG CGC GTT GAA
E M L E L E M D M E A D L G I D S I K R V E>

18440 18460 18480 18500
ATT CTT GGC ACA GTA CAA GAT GAG CTA CCG GGT TTA CCT GAG CTA AAT CCA GAA GAT TTG GCA GAG
I L G T V Q D E L P G L P E L N P E D L A E>

18520 18540 18560
TGT CGT ACT CTT GGC GAA ATC GTG ACT TAT ATG AAC TCT AAA CTC GCT GAC GGC TCT AAG CTG CCA
C R T L G E I V T Y M N S K L A D G S K L P>

18580 18600 18620
GCT GAA GGC TCT ATG CAC TAT CAG CTG TCT ACA AGT ACC GCT GCT GCG ACT CCT GTA GCG AAT GGT
A E G S M H Y Q L S T S T A A A T P V A N G>

18640 18660 18680
CTC TCT GCA GAA AAA GTT CAA GCG ACC ATG ATG TCT GTA GTT GCA GAT AAA ACT GGC TAC CCA ACT
L S A E K V Q A T M M S V V A D K T G Y P T>

18700 18720 18740 18760
GAA ATG CTT GAA CTT GAA ATG GAT ATG GAA GCC GAT TTA GGT ATC GAT TCT ATC AAG CGC GTT GAA
E M L E L E M D M E A D L G I D S I K R V E>

18780 18800 18820
ATT CTT GGC ACA GTA CAA GAT GAG CTA CCG GGT TTA CCT GAG CTA AAT CCA GAA GAT CTA GCA GAG
I L G T V Q D E L P G L P E L N P E D L A E>

18840 18860 18880

Fig. 4
14/430

MISSING AT THE TIME OF PUBLICATION

20020 20040 20060 20080
GTT AGC AAT GCG TTC TTG TGG GCC AAA TTA TTG CAA CCA AAG CTC GTT GCT GGA GCA GAT GCG CGT
V S N A F L W A K L Q P K L V A G A D A R>
20100 20120 20140
CGC TGT TTT GTA ACA GTA AGC CGT ATC GAC GGT GGC TTA GGT TAC CTA AAT ACT GAC GCC CTA AAA
R C F V T V S R I D G G F G Y L N T D A L K>
20160 20180 20200
GAT GCT GAG CTA AAC CAA GCA GCA TTA GCT GGT TTA ACT AAA ACC TTA AGC CAT GAA TGG CCA CAA
D A E L N Q A A L A G L T L S H E W P Q>
20220 20240 20260 20280
GTG TTC TGT CGC GCG CTA GAT ATT GCA ACA GAT GTT GAT GCA ACC CAT CTT GCT GAT GCA ATC ACC
V F C R A L D I A T D V D A T H L A D A I T>
20300 20320 20340
AGT GAA CTA TTT GAT AGC CAA GCT CAG CTA CCT GAA GTG GGC TTA AGC TTA ATT GAT GGC AAA GTT
S E L F D S Q A Q L P E V G L S L I D G K V>
20360 20380 20400
AAC CGC GTA ACT CTA GTT GCT GCT GAA GCT GCA GAT AAA ACA GCA AAA GCA GAG CTT AAC AGC ACA
N R V T L V A A E A A D K T A K A E L N S T>
20420 20440 20460 20480
GAT AAA ATC TTA GTG ACT GGT GGG GCA AAA GGG GTG ACA TTT GAA TGT GCA CTG GCA TTA GCA TCT
D K I L V T G G A K G V T F E C A L A L A S>
20500 20520 20540
CGC AGC CAG TCT CAC TTT ATC TTA GCT GGG CGC AGT GAA TTA CAA GCT TTA CCA AGC TGG GCT GAG
R S Q S H F I L A G R S E L Q A L P S W A E>
20560 20580 20600
GGT AAG CAA ACT AGC GAG CTA AAA TCA GCT GCA ATC GCA CAT ATT ATT TCT ACT GGT CAA AAG CCA
G K Q T S E L S A A I A H I I S T G Q K P>
20620 20640 20660
ACG CCT AAG CAA GTT GAA GCC GCT GTG TGG CCA GTG CAA AGC AGC ATT GAA ATT AAT GCC GCC CTA
T P K Q V E A A V W P V Q S S I E I N A A L>
20680 20700 20720 20740
GCC GCC TTT AAC AAA GTT GGC GCC TCA GCT GAA TAC GTC AGC ATG GAT GTT ACC GAT AGC GCC CCA
A A F N K V G A S A E Y V S M D V T D S A A>
20760 20780 20800
ATC ACA GCA GCA CTT AAT GGT CGC TCA AAT GAG ATC ACC GGT CTT ATT CAT GGC GCA GGT GTA CTA
I T A A L N G R S N E I T G L I H G A G V L>
20820 20840 20860
GCC GAC AAG CAT ATT CAA GAC AAG ACT CTT GCT GAA CTT GCT AAA GTT TAT GGC ACT AAA GTC AAC
A D K H I Q D K T L A E L A K V Y G T K V N>
20880 20900 20920 20940
GGC CTA AAA GCG CTG CTC GCG GCA CTT GAG CCA AGC AAA ATT AAA TTA CTT GCT ATG TTC TCA TCT
G L K A L L A A L E P S K I K L L A M F S S>
20960 20980 21000
GCA GCA GGT TTT TAC GGT AAT ATC GGC CAA AGC GAT TAC GCG ATG TCG AAC GAT ATT CTT AAC AAG
A A G F Y G N I G Q S D Y A M S N D I L N K>
21020 21040 21060
GCA GCG CTG CAG TTC ACC GCT CGC AAC CCA CAA GCT AAA GTC ATG AGC TTT AAC TGG GGT CCT TGG
A A L Q F T A R N P Q A K V M S F N W G P W>
21080 21100 21120 21140
GAT GGC GGC ATG GTT AAC CCA GCG CTT AAA AAG ATG TTT ACC GAG CGT GGT GTG TAC GTT ATT CCA
D G G M V N P A L K K M F T E R G V Y V I P>
21160 21180 21200
CTA AAA GCA GGT GCA GAG CTA TTT GCC ACT CAG CTA TTG GCT GAA ACT GGC GTG CAG TTG CTC ATT
L K A G A E L F A T Q L L A E T G V Q L L I>

Fig. 4
16/30

21220 21240 21260
GGT ACG TCA ATG CAA GGT GGC AGC GAC ACT AAA GCA ACT GAG ACT GCT TCT GTA AAA AAG CTT AAT
G T S M Q G G S D T K A T E T A S V K K L N>
21280 21300 21320
GCG GGT GAG GTG CTA AGT GCA TCG CAT CCG CGT GCT GGT GCA CAA AAA ACA CCA CTA CAA GCT GTC
A G E V L S A S H P R A G A Q K T P L Q A V>
21340 21360 21380 21400
ACT GCA ACG CGT CTG TTA ACC CCA AGT GCC ATG GTC TTC ATT GAA GAT CAC CGC ATT GGC GGT AAC
T A T R L L T P S A M V F I E D H R I G G N>
21420 21440 21460
AGT GTG TTG CCA ACG GTA TGC GCC ATC GAC TGG ATG CGT GAA GCG GCA AGC GAC ATG CTT GGC GCT
S V L P T V C A I D W M R E A A S D M L G A>
21480 21500 21520
CAA GTT AAG GTA CTT GAT TAC AAG CTA TTA AAA GGC ATT GTA TTT GAG ACT GAT GAG CCG CAA GAG
Q V K V L D Y K L L K G I V F E T D E P Q E>
21540 21560 21580 21600
TTA ACA CTT GAG CTA ACG CCA GAC GAT TCA GAC GAA GCT ACG CTA CAA GCA TTA ATC AGC TGT AAT
L T L E L T P D D S D E A T L Q A L I S C N>
21620 21640 21660
GGG CGT CCG CAA TAC AAG GCG ACG CTT ATC AGT GAT AAT GCC GAT ATT AAG CAA CTT AAC AAG CAG
G R P Q Y K A T L I S D N A D I K Q L N K Q>
21680 21700 21720
TTT GAT TTA AGC GCT AAG GCG ATT ACC ACA GCA AAA GAG CTT TAT AGC AAC GGC ACC TTG TTC CAC
F D L S A K A I T T A K E L Y S N G T L F H>
21740 21760 21780 21800
GGT CCG CGT CTA CAA GGG ATC CAA TCT GTA GTG CAG TTC GAT GAT CAA GGC TTA ATT GCT AAA GTC
G P R L Q G I Q S V V Q F D D Q G L I A K V>
21820 21840 21860
GCT CTG CCT AAG GTT GAA CTT AGC GAT TGT GGT GAG TTC TTG CCG CAA ACC CAC ATG GGT GGC AGT
A L P K V E L S D C G E F L P Q T H M G G S>
21880 21900 21920
CAA CCT TTT GCT GAG GAC TTG CTA TTA CAA GCT ATG CTG GTT TGG GCT CGC CTT AAA ACT GGC TCG
Q P F A E D L L Q A M L V W A R L K T G S>
21940 21960 21980
GCA AGT TTG CCA TCA AGC ATT GGT GAG TTT ACC TCA TAC CAA CCA ATG GCC TTT GGT GAA ACT GGT
A S L P S S I G E F T S Y Q P M A F G E T G>
22000 22020 22040 22060
ACC ATA GAG CTT GAA GTG ATT AAG CAC AAC AAA CGC TCA CTT GAA GCG AAT GTT GCG CTA TAT CGT
T I E L E V I K H N K R S L E A N V A L Y R>
22080 22100 22120
GAC AAC GGC GAG TTA AGT GCC ATG TTT AAG TCA GCT AAA ATC ACC ATT AGC AAA AGC TTA AAT TCA
D N G E L S A M F K S A K I T I S K S L N S>
22140 22160 22180 22200
GCA TTT TTA CCT GCT GTC TTA GCA AAC GAC AGT GAG GCG AAT TAGTGGA ACAACGCGCT AAAGCTAGTG
A F L P A V L A N D S E A N>
22220 22240 22260
CG ATG CCG CTG CGC ATC GCA CTT ATC TTA CTG CCA ACA CCG CAG TTT GAA GTT AAC TCT GTC GAC
M P L R I A L I L L P T P Q F E V N S V D>
22280 22300 22320
CAG TCA GTA TTA GCC AGC TAT CAA ACA CTG CAG CCT GAG CTA AAT GCC CTG CTT AAT AGT GCG CCG
Q S V L A S Y Q T L Q P E L N A L L N S A P>
22340 22360 22380
ACA CCT GAA ATG CTC AGC ATC ACT ATC TCA GAT GAT AGC GAT GCA AAC AGC TTT GAG TCG CAG CTA

o-f 7

Fig. 4

17/30

T P E M L S I T I S D D S D A N S F E S Q L>

22400 22420 22440 22460
AAT GCT GCG ACC AAC GCA ATT AAC AAT GGC TAT ATC GTC AAG CTT GCT ACG GCA ACT CAC GCT TTG
N A A T N A I N N G Y I V K L A T A T H A L>

22480 22500 22520
TTA ATG CTG CCT GCA TTA AAA GCG GCG CAA ATG CGG ATC CAT CCT CAT GCG CAG CTT GCC GCT ATG
L M L P A L K A A Q M R I H P H A Q L A A M>

22540 22560 22580
CAG CAA GCT AAA TCG ACG CCA ATG AGT CAA GTA TCT GGT GAG CTA AAG CTT GGC GCT AAT GCG CTA
Q Q A K S T P M S Q V S G E L K L G A N A L>

22600 22620 22640 22660
AGC CTA GCT CAG ACT AAT GCG CTG TCT CAT GCT TTA AGC CAA GCC AAG CGT AAC TTA ACT GAT GTC
S L A Q T N A L S H A L S Q A K R N L T D V>

22680 22700 22720
AGC GTG AAT GAG TGT TTT GAG AAC CTC AAA AGT GAA CAG CAG TTC ACA GAG GTT TAT TCG CTT ATT
S V N E C F E N L K S E Q Q F T E V Y S L I>

22740 22760 22780
CAG CAA CTT GCT AGC CGC ACC CAT GTG AGA AAA GAG GTT AAT CAA GGT GTG GAA CTT GGC CCT AAA
Q Q L A S R T H V R K E V N Q G V E L G P K>

22800 22820 22840
CAA GCC AAA AGC CAC TAT TGG TTT AGC GAA TTT CAC CAA AAC CGT GTT GCT GCC ATC AAC TTT ATT
Q A K S H Y W F S E F H Q N R V A A I N F I>

22860 22880 22900 22920
AAT GGC CAA CAA GCA ACC AGC TAT GTG CTT ACT CAA GGT TCA GGA TTG TTA GCT GCG AAA TCA ATG
N G Q Q A T S Y V L T Q G S G L L A A K S M>

22940 22960 22980
CTA AAC CAG CAA AGA TTA ATG TTT ATC TTG CCG GGT AAC AGT CAG CAA CAA ATA ACC GCA TCA ATA
L N Q Q R L M F I L P G N S Q Q Q I T A S I>

23000 23020 23040
ACT CAG TTA ATG CAG CAA TTA GAG CGT TTG CAG GTA ACT GAG GTT AAT GAG CTT TCT CTA GAA TGC
T Q L M Q Q L E R L Q V T E V N E L S L E C>

23060 23080 23100 23120
CAA CTA GAG CTG CTC AGC ATA ATG TAT GAC AAC TTA GTC AAC GCA GAC AAA CTC ACT ACT CGC GAT
Q L E L L S I M Y D N L V N A D K L T T R D>

23140 23160 23180
AGT AAG CCC GCT TAT CAG GCT GTG ATT CAA GCA AGC TCT GTT AGC GCT GCA AAG CAA GAG TTA AGC
S K P A Y Q A V I Q A S S V S A A K Q E L S>

23200 23220 23240
GCG CTT AAC GAT GCA CTC ACA GCG CTG TTT GCT GAG CAA ACA AAC GCC ACA TCA ACG AAT AAA GGC
A L N D A L T A L F A E Q T N A T S T H K G>

23260 23280 23300 23320
TTA ATC CAA TAC AAA ACA CCG GCG GGC AGT TAC TTA ACC CTA ACA CCG CTT GCG AGC AAC AAT GAC
L I Q Y K T P A G S Y L T L T P L G S N N D>

23340 23360 23380
AAC GCC CAA GCG GGT CTT GCT TTT GTC TAT CCG GGT GTG GGA ACG GTT TAC GCC GAT ATG CTT AAT
N A Q A G L A F V Y P G V G T V Y A D M L N>

23400 23420 23440
GAG CTG CAT CAG TAC TTC CCT GCG CTT TAC GCC AAA CTT GAG CGT GAA GGC GAT TTA AAG GCG ATG
E L H Q Y F P A L Y A K L E R E G D L K A M>

23460 23480 23500
CTA CAA GCA GAA GAT ATC TAT CAT CTT GAC CCT AAA CAT GCT GCC CAA ATG AGC TTA GGT GAC TTA
L Q A E D I Y H L D P K H A A Q M S L G D L>

23520 23540 23560 23580

Fig. 4
18/30

GCC ATT GCT GGC GTG GGG AGC AGC TAC CTG TTA ACT CAG CTG CTC ACC GAT GAG TTT AAT ATT AAG
A I A G V G S S Y L L T Q L L T D E F N I K>

23600 23620 23640
CCT AAT TTT GCA TTA GGT TAC TCA ATG GGT GAA GCA TCA ATG TGG GCA AGC TTA GGC GTA TGG CAA
P N F A L G Y S M G E A S M W A S L G V W Q>

23660 23680 23700
AAC CCG CAT GCG CTG ATC AGC AAA ACC CAA ACC GAC CCG CTA TTT ACT TCT GCT ATT TCC GGC AAA
N P H A L I S K T Q T D P L F T S A I S G K>

23720 23740 23760 23780
TTG ACC GCG GTT AGA CAA GCT TGG CAG CTT GAT GAT ACC GCA GCG GAA ATC CAG TGG AAT AGC TTT
L T A V R Q A W Q L D D T A A E I Q W N S F>

23800 23820 23840
GTG GTT AGA AGT GAA GCA GCG CCG ATT GAA GCC TTG CTA AAA GAT TAC CCA CAC GCT TAC CTC GCG
V V R S E A A P I E A L L K D Y P H A Y L A>

23860 23880 23900
ATT ATT CAA GGG GAT ACC TGC GTA ATC GCT GGC TGT GAA ATC CAA TGT AAA GCG CTA CTT GCA GCA
I I Q G D T C V I A G C E I Q C K A L L A A>

23920 23940 23960 23980
CTG GGT AAA CGC GGT ATT GCA GCT AAT CGT GTA ACG GCG ATG CAT ACG CAG CCT GCG ATG CAA GAG
L G K R G I A A N R V T A M H T Q P A M Q E>

24000 24020 24040
CAT CAA AAT GTG ATG GAT TTT TAT CTG CAA CCG TTA AAA GCA GAG CTT CCT AGT GAA ATA AGC TTT
H Q N V M D F Y L Q P L K A E L P S E I S F>

24060 24080 24100
ATC AGC GCC GCT GAT TTA ACT GCC AAG CAA ACG GTG AGT GAG CAA GCA CTT AGC AGC CAA GTC GTT
I S A A D L T A K Q T V S E Q A L S S Q V V>

24120 24140 24160
GCT CAG TCT ATT GCC GAC ACC TTC TGC CAA ACC TTG GAC TTT ACC GCG CTA GTA CAT CAC GCC CAA
A Q S I A D T F C Q T L D F T A L V H H A Q>

24180 24200 24220 24240
CAT CAA GGC GCT AAG CTG TTT GTT GAA ATT GGC GCG GAT AGA CAA AAC TGC ACC TTG ATA GAC AAG
H Q G A K L F V E I G A D R Q N C T L I D K>

24260 24280 24300
ATT GTT AAA CAA GAT GGT GCC AGC AGT GTA CAA CAT CAA CCT TGT TGC ACA GTG CCT ATG AAC GCA
I V K Q D G A S S V Q H Q P C C T V P M N A>

24320 24340 24360
AAA GGT AGC CAA GAT ATT ACC AGC GTG ATT AAA GCG CTT GGC CAA TTA ATT AGC CAT CAG GTG CCA
K G S Q D I T S V I K A L G Q L I S H Q V P>

24380 24400 24420 24440
TTA TCG GTG CAA CCA TTT ATT GAT GGA CTC AAG CGC GAG CTA ACA CTT TGC CAA TTG ACC AGC CAA
L S V Q P F I D G L K R E L T L C Q L T S Q>

24460 24480 24500
CAG CTG GCA GCA CAT GCA AAT GTT GAC AGC AAG TTT CAG TCT AAC CAA GAC CAT TTA CTT CAA GGG
Q L A A H A N V D S K F E S N Q D H L L Q G>

24520 24540 24560
GAA GTC TA ATG TCA TTA CCA GAC AAT GCT TCT AAC CAC CTT TCT GCC AAC CAG AAA GGC GCA TCT
E V>

24580 24600 24620 24640
CAG GCA AGT AAA ACC AGT AAG CAA AGC AAA ATC GCC ATT GTC GGT TTA GCC ACT CTG TAT CCA GAC
Q A S K T S K Q S K I A I V G L A T L Y P D>

24660 24680 24700
GCT AAA ACC CCG CAA GAA TTT TGG CAG AAT TTG CTG GAT AAA CGC GAC TCT CGC AGC ACC TTA ACT
A K T P Q E F W Q N L L D K R D S R S T L T>

5-8

Fig. 4
19/30

24720 AAC GAA AAA CTC GGC GCT AAC AGC CAA GAT TAT CAA GGT GTG CAA GGC CAA TCT GAC CGT TTT TAT
N E K L G A N S Q D Y Q G V Q G Q S D R F Y>

24780 TGT AAT AAA GGC GGC TAC ATT GAG AAC TTC AGC TTT AAT GCT GCA GGC TAC AAA TTG CCG GAG CAA
C N K G G Y I E N F S F N A A G Y K L P E Q>

24840 AGC TTA AAT GGC TTG GAC GAC AGC TTC CTT TGG GCG CTC GAT ACT AGC CGT AAC GCA CTA ATT GAT
S L N G L D D S F L W A L D T S R N A L I D>

24920 GCT GGT ATT GAT ATC AAC GGC GCT GAT TTA AGC CGC GCA GGT GTA GTC ATG GGC GCG CTG TCG TTC
A G I D I N G A D L S R A G V V M G A L S F>

24980 CCA ACT ACC CGC TCA AAC GAT CTG TTT TTG CCA ATT TAT CAC AGC GCC GTT GAA AAA GCC CTG CAA
P T T R S N D L F L P I Y H S A V E K A L Q>

25040 GAT AAA CTA GGC GTA AAG GCA TTT AAG CTA AGC CCA ACT AAT GCT CAT ACC GCT CGC GCG GCA AAT
D K L G V K A F K L S P T N A H T A R A A N>

25120 GAG AGC AGC CTA AAT GCA GCC AAT GGT GCC ATT GCC CAT AAC AGC TCA AAA GTG GTG GCC GAT GCA
E S S L N A A N G A I A H N S K V V A D A>

25180 CTT GGC CTT GGC GGC GCA CAA CTA AGC CTA GAT GCT GCC TGT GCT AGT TCG GTT TAC TCA TTA AAG
L G L G G A Q L S L D A A C S S V Y S L K>

25240 CTT GCC TGC GAT TAC CTA AGC ACT GGC AAA GCC GAT ATC ATG CTA GCA GGC GCA GTA TCT GGC GCG
L A C D Y L S T G K A D I M L A G A V S G A>

25320 GAT CCT TTC TTT ATT AAT ATG GGA TTC TCA ATC TTC CAC GCC TAC CCA GAC CAT GGT ATC TCA GTA
D P F F I N M G F S I F H A Y P D H G I S V>

25380 CCG TTT GAT GCC AGC AGT AAA GGT TTG TTT GCT GGC GAA GGC GCT GGC GTA TTA GTG CTT AAA CGT
P F D A S S K G L F A G E G A G V L V L K R>

25440 CTT GAA GAT GCC GAG CGC GAC AAT GAC AAA ATC TAT GCG GTT GTT AGC GGC GTA GGT CTA TCA AAC
L E D A E R D N D K I Y A V V S G V G L S N>

25500 GAC GGT AAA GGC CAG TTT GTA TTA AGC CCT AAT CCA AAA GGT CAG GTG AAG GCC TTT GAA CGT GCT
D G K G Q F V L S P N P K G Q V K A F E R A>

25580 TAT GCT GCC AGT GAC ATT GAG CCA AAA GAC ATT GAA GTG ATT GAG TGC CAC GCA ACA GGC ACA CCG
Y A A S D I E P K D I E V I E C H A T G T P>

25640 CTT GGC GAT AAA ATT GAG CTC ACT TCA ATG GAA ACC TTC TTT GAA GAC AAG CTG CAA GGC ACC GAT
L G D K I E L S M E T F E D K L Q G T D>

25700 GCA CCG TTA ATT GGC TCA GCT AAG TCT AAC TTA GGC CAC CTA TTA ACT GCA GCG CAT GCG GGG ATC
A P L I G S A K S N L G H L L T A A H A G I>

25780 ATG AAG ATG ATC TTC GCC ATG AAA GAA GGT TAC CTG CCG CCA AGT ATC AAT ATT AGT GAT GCT ATC
M K M I F A M K E G Y L P P S I N I S D A I>

25840 GCT TCG CCG AAA AAA CTC TTC GGT AAA CCA ACC CTG CCT AGC ATG GTT CAA GGC TGG CCA GAT AAG
A S P K K L F G K P T L P S M V Q G W P D K>

Fig. 4
20/30

KAS Tam

25900 25920 25940 25960
CCA TCG AAT AAT CAT TTT GGT GTA AGA ACC CGT CAC GCA GGC GTA TCG GTA TTT GGC TTT GGT GGC
P S N N H F G V R T R H A G V S V F G F G G>

25980 26000 26020
TGT AAC GCC CAT CTG TTG CTT GAG TCA TAC AAC GGC AAA GGA ACA GTA AAG GCA GAA GCC ACT CAA
C N A H L L L E S Y N G K G T V K A E A T O>

26040 26060 26080
GTA CCG CGT CAA GCT GAG CCG CTA AAA GTG GTT GGC CTT GCC TCG CAC TTT GGG CCT CTT AGC AGC
V P R Q A E P L K V V G L A S H F G P L S S>

26100 26120 26140
ATT AAT GCA CTC AAC AAT GCT GTG ACC CAA GAT GGG AAT GGC TTT ATC GAA CTG CCG AAA AAG CGC
I N A L N N A V T Q D G N G F I E L P K K R>

26160 26180 26200 26220
TGG AAA GGC CTT GAA AAG CAC AGT GAA CTG TTA GCT GAA TTT GGC TTA GCA TCT GCG CCA AAA GGT
W K G L E K H S E L L A E F G L A S A P K G>

26240 26260 26280
GCT TAT GTT GAT AAC TTC GAG CTG GAC TTT TTA CGC TTT AAA CTG CCG CCA AAC GAA GAT GAC CGT
A Y V D N F E L D F L R F K L P P N E D D R>

26300 26320 26340
TTG ATC TCA CAG CAG CTA ATG CTA ATG CGA GTA ACA GAC GAA GCC ATT CGT GAT GCC AAG CTT GAG
L I S Q Q L M L M R V T D E A I R D A K L E>

26360 26380 26400 26420
CCG GGG CAA AAA GTA GCT GTA TTA GTG GCA ATG GAA ACT GAG CTT GAA CTG CAT CAG TTC CGC GGC
P G Q K V A V L V A M E T E L E L H Q F R G>

26440 26460 26480
CGG GTT AAC TTG CAT ACT CAA TTA GCG CAA AGT CTT GCC GCC ATG GGC GTG AGT TTA TCA ACG GAT
R V N L H T Q L A Q S L A A M G V S L S T D>

26500 26520 26540
GAA TAC CAA GCG CTT GAA GCC ATC GCC ATG GAC AGC GTG CTT GAT GCT GCC AAG CTC AAT CAG TAC
E Y Q A L E A I A M D S V L D A A K L N Q Y>

26560 26580 26600 26620
ACC AGC TTT ATT GGT AAT ATT ATG GCG TCA CGC GTG GCG TCA CTA TGG GAC TTT AAT GGC CCA GCC
T S F I G N I M A S R V A S L W D F N G P A>

26640 26660 26680
TTC ACT ATT TCA GCA GCA GAG CAA TCT GTG AGC CGC TGT ATC GAT GTG GCG CAA AAC CTC ATC ATG
F T I S A A E Q S V S R C I D V A Q N L I M>

26700 26720 26740
GAG GAT AAC CTA GAT GCG GTG GTG ATT GCA GCG GTC GAT CTC TCT GGT AGC TTT GAG CAA GTC ATT
E D N L D A V V I A A V D L S G S F E Q V I>

26760 26780 26800
CTT AAA AAT GCC ATT GCA CCT GTA GCC ATT GAG CCA AAC CTC GAA GCA AGC CTT AAT CCA ACA TCA
L K N A I A P V A I E P N L E A S L N P T S>

26820 26840 26860 26880
GCA AGC TGG AAT GTC GGT GAA GGT GCT GGC GCG GTC GTG CTT GTT AAA AAT GAA GCT ACA TCG GGC
A S W N V G E G A G A V V L V K N E A T S O>

26900 26920 26940
TGC TCA TAC GGC CAA ATT GAT GCA CTT GGC TTT GCT AAA ACT GCC GAA ACA GCG TTG GCT ACC GAC
C S Y G Q I D A L G F A K T A E T A L A T D>

26960 26980 27000
AAG CTA CTG AGC CAA ACT GCC ACA GAC TTT AAT AAG GTT AAA GTG ATT GAA ACT ATG GCA GCG CCT
K L L S Q T A T D F N K V K V I E T M A A P>

27020 27040 27060 27080
GCT AGC CAA ATT CAA TTA GCG CCA ATA GTT AGC TCT CAA GTG ACT CAC ACT GCT GCA GAG CAG CGT

Fig. 4
21/30

A S Q I Q L A P I V S S Q V T H T A A E Q R>
27100 27120 27140
GTT GGT CAC TGC TTT GCT GCA GCG GGT ATG GCA AGC CTA TTA CAC GGC TTA CTT AAC TTA AAT ACT
V G H C F A A A G M A S L L H G L L N L N T>
27160 27180 27200
GTA GCC CAA ACC AAT AAA GCC AAT TGC GCG CTT ATC AAC AAT ATC AGT GAA AAC CAA TTA TCA CAG
V A Q T N K A N C A L I N N I S E N Q L S Q>
27220 27240 27260 27280
CTG TTG ATT AGC CAA ACA GCG AGC GAA CAA CAA GCA TTA ACC GCG CGT TTA AGC AAT GAG CTT AAA
L L I S Q T A S E Q Q A L T A R L S N E L K>
27300 27320 27340
TCC GAT GCT AAA CAC CAA CTG GTT AAG CAA GTC ACC TTA GGT GGC CGT GAT ATC TAC CAG CAT ATT
S D A K H Q L V K Q V T L G G R D I Y Q H I>
27360 27380 27400
GTT GAT ACA CCG CTT GCA AGC CTT GAA AGC ATT ACT CAG AAA TTG GCG CAA GCG ACA GCA TCG ACA
V D T P L A S L E S I T Q K L A Q A T A S T>
27420 27440 27460
GTG GTC AAC CAA GTT AAA CCT ATT AAG GCC GCT GGC TCA GTC GAA ATG GCT AAC TCA TTC GAA ACG
V V N Q V K P I K A A G S V E M A N S F E T>
27480 27500 27520 27540
GAA AGC TCA GCA GAG CCA CAA ATA ACA ATT GCA GCA CAA CAG ACT GCA AAC ATT GGC GTC ACC GCT
E S S A E P Q I T I A A Q Q T A N I G V T A>
27560 27580 27600
CAG GCA ACC AAA CGT GAA TTA GGT ACC CCA CCA ATG ACA ACA AAT ACC ATT GCT AAT ACA GCA AAT
Q A T K R E L G T P P M T T N T I A N T A N>
27620 27640 27660
AAT TTA GAC AAG ACT CTT GAG ACT GTT GCT GGC AAT ACT GTT GCT AGC AAG GTT GGC TCT GGC GAC
N L D K T L E T V A G N T V A S K V G S G D>
27680 27700 27720 27740
ATA GTC AAT TTT CAA CAG AAC CAA CAA TTG GCT CAA CAA GCT CAC CTC GCC TTT CTT GAA AGC CGC
I V N F Q Q N Q Q L A Q Q A H L A F L E S R>
27760 27780 27800
AGT GCG GGT ATG AAG GTG GCT GAT GCT TTA TTG AAG CAA CAG CTA GCT CAA GTA ACA GGC CAA ACT
S A G M K V A D A L L K Q Q L A Q V T G Q T>
27820 27840 27860
ATC GAT AAT CAG GCC CTC GAT ACT CAA GCC GTC GAT ACT CAA ACA AGC GAG AAT GTA GCG ATT GCC
I D N Q A L D T Q A V D T Q T S E N V A I A>
27880 27900 27920 27940
GCA GAA TCA CCA GTT CAA GTT ACA ACA CCT GTT CAA GTT ACA ACA CCT GTT CAA ATC AGT GTT GTG
A E S P V Q V T T P V Q V T T P V Q I S V V>
27960 27980 28000
GAG TTA AAA CCA GAT CAC GCT AAT GTG CCA CCA TAC ACG CCG CCA GTG CCT GCA TTA AAG CCG TGT
E L K P D H A N V P P Y T P P V P A L K P C>
28020 28040 28060
ATC TGG AAC TAT GCC GAT TTA GTT GAG TAC GCA GAA GGC GAT ATC GCC AAG GTA TTT GGC AGT GAT
I W N Y A D L V E Y A E G D I A K V F G S D>
28080 28100 28120
TAT GCC ATT ATC GAC AGC TAC TCG CGC CGC GTA CGT CTA CCG ACC ACT GAC TAC CTG TTG GTA TCG
Y A I I D S Y S R R V R L P T T D Y L L V S>
28140 28160 28180 28200
CGC GTG ACC AAA CTT GAT GCG ACC ATC AAT CAA TTT AAG CCA TGC TCA ATG ACC ACT GAG TAC GAC
R V T K L D A T I N Q P K P C S M T T E Y D>
28220 28240 28260

Fig. 4
22/30

ATC CCT GTT GAT GCG CCG TAC TTA GTA GAC GGA CAA ATC CCT TGG GCG GTA GCA GTA GAA TCA GGC
I P V D A P Y L V D G Q I P W A V A V E S G>

28280 28300 28320
CAA TGT GAC TTG ATG CTT ATT AGC TAT CTC GGT ATC GAC TTT GAG AAC AAA GGC GAG CCG GTT TAT
Q C D L M L I S Y L G I D F E N K G E R V Y>

28340 28360 28380 28400
CGA CTA CTC GAT TGT ACC CTC ACC TTC CTA GGC GAC TTG CCA CGT GGC GGA GAT ACC CTA CGT TAC
R L L D C T L T F L G D L P R G G D T L R Y>

28420 28440 28460
GAC ATT AAG ATC AAT AAC TAT GCT CGC AAC GGC GAC ACC CTG CTG TTC TTC TTC TCG TAT GAG TGT
D I K I N N Y A R N G D T L L F F F S Y E C>

28480 28500 28520
TTT GTT GGC GAC AAG ATG ATC CTC AAG ATG GAT GGC GGC TGC GCT GGC TTC TTC ACT GAT GAA GAG
F V G D K M I L K M D G G C A G F F T D E E>

28540 28560 28580 28600
CTT GCC GAC GGT AAA GGC GTG ATT CGC ACA GAA GAA GAG ATT AAA GCT CGC AGC CTA GTG CAA AAG
L A D G K G V I R T E E E I K A R S L V Q K>

28620 28640 28660
CAA CGC TTT AAT CCG TTA CTA GAT TGT CCT AAA ACC CAA TTT AGT TAT GGT GAT ATT CAT AAG CTA
Q R F N P L L D C P K T Q F S Y G D I H K L>

28680 28700 28720
TTA ACT GCT GAT ATT GAG GGT TGT TTT GGC CCA AGC CAC AGT GGC GTC CAC CAG CCG TCA CTT TGT
L T A D I E G C F G P S H S G V H Q P S L C>

28740 28760 28780
TTC GCA TCT GAA AAA TTC TTG ATG ATT GAA CAA GTC AGC AAG GTT GAT CGC ACT GGC GGT ACT TGG
F A S E K F L M I E Q V S K V D R T G G T W>

28800 28820 28840 28860
GGA CTT GGC TTA ATT GAG GGT CAT AAG CAG CTT GAA GCA GAC CAC TGG TAC TTC CCA TGT CAT TTC
G L G L I E G H K Q L E A D H W Y F P C H F>

28880 28900 28920
AAG GGC GAC CAA GTG ATG GCT GGC TCG CTA ATG GCT GAA GGT TGT GGC CAG TTA TTG CAG TTC TAT
K G D Q V M A G S L M A E G C G Q L L Q F Y>

28940 28960 28980
ATG CTG CAC CTT GGT ATG CAT ACC CAA ACT AAA AAT GGT CGT TTC CAA CCT CTT GAA AAC GCC TCA
M L H L G M H T Q T K N G R F Q P L E N A S>

29000 29020 29040 29060
CAG CAA GTA CCG TGT CCG GGT CAA GTG CTG CCA CAA TCA GGC GTG CTA ACT TAC CGT ATG GAA GTG
Q Q V R C R G Q V L P Q S G V L T Y R M E V>

29080 29100 29120
ACT GAA ATC GGT TTC AGT CCA CGC CCA TAT GCT AAA GCT AAC ATC GAT ATC TTG CTT AAT GGC AAA
T E I G F S P R P Y A K A N I D I L L N G K>

29140 29160 29180
GCG GTA GTG GAT TTC CAA AAC CTA GGC GTG ATG ATA AAA GAG GAA GAT GAG TGT ACT CGT TAT CCA
A V V D F Q N L G V M I K E E D E C T R Y P>

29200 29220 29240 29260
CTT TTG ACT GAA TCA ACA ACG GCT AGC ACT GCA CAA GTA AAC GCT CAA ACA AGT GCG AAA AAG GTA
L L T E S T T A S T A Q V N A Q T S A K K V>

29280 29300 29320
TAC AAG CCA GCA TCA GTC AAT GCG CCA TTA ATG GCA CAA ATT CCT GAT CTG ACT AAA GAG CCA AAC
Y K P A S V N A P L M A Q I P D L T K E P N>

29340 29360 29380
AAG GGC GTT ATT CCG ATT TCC CAT GTT GAA GCA CCA ATT ACG CCA GAC TAC CCG AAC CGT GTA CCT
K G V I P I S H V E A P I T P D Y P N R V P>

29400 29420 29440

Fig. 4
23/30

GAT ACA GTG CCA TTC ACC CCG TAT CAC ATG TTT GAG TTT GCT ACA GGC AAT ATC GAA AAC TGT TTC
D T V P F T P Y H M F E F A T G N E N C F>

29460 9480 29500 29520
GGG CCA GAG TTC TCA ATC TAT CGC GGC ATG ATC CCA CCA CGT ACA CCA TGC GGT GAC TTA CAA GTG
G P E F S I Y R G M I P P R T P C G D L Q V>

29540 29560 29580
ACC ACA CGT GTG ATT GAA GTT AAC GGT AAG CGT GGC GAC TTT AAA AAG CCA TCA TCG TGT ATC GCT
T T R V I E V N G K R G D F K K P S S C I A>

29600 29620 29640
GAA TAT GAA GTG CCT GCA GAT GCG TGG TAT TTC GAT AAA AAC AGC CAC GGC GCA GTG ATG CCA TAT
E Y E V P A D A W Y F D K N S H G A V M P Y>

29660 29680 29700 29720
TCA ATT TTA ATG GAG ATC TCA CTG CAA CCT AAC GGC TTT ATC TCA GGT TAC ATG GGC ACA ACC CTA
S I L M E I S L Q P N G F I S G Y M G T T L>

29740 29760 29780
GGC TTC CCT GGC CTT GAG CTG TTC TTC CGT AAC TTA GAC GGT AGC GGT GAG TTA CTA CGT GAA GTA
G F P G L E L F F R N L D G S G E L L R E V>

29800 29820 29840
GAT TTA CGT GGT AAA ACC ATC CGT AAC GAC TCA CGT TTA TTA TCA ACA GTG ATG GCC GGC ACT AAC
D L R G K T I R N D S R L L S T V M A G T N>

29860 29880 29900 29920
ATC ATC CAA AGC TTT AGC TTC GAG CTA AGC ACT GAC GGT GAG CCT TTC TAT CGC GGC ACT GCG GTA
I I Q S F S F E L S T D G E P F Y R G T A V>

29940 29960 29980
TTT GGC TAT TTT AAA GGT GAC GCA CTT AAA GAT CAG CTA GGC CTA GAT AAC GGT AAA GTC ACT CAG
F G Y F K G D A L K D Q L G L D N G K V T Q>

30000 30020 30040
CCA TGG CAT GTA GCT AAC GGC GTT GCT GCA AGC ACT AAG GTG AAC CTG CTT GAT AAG AGC TGC CGT
P W H V A N G V A A S T K V N L L D K S C R>

30060 30080 30100
CAC TTT AAT GCG CCA GCT AAC CAG CCA CAC TAT CGT CTA GCC GGT GGT CAG CTG AAC TTT ATC GAC
H F N A P A N Q P H Y R L A G G Q L N F I D>

30120 30140 30160 30180
AGT GTT GAA ATT GTT GAT AAT GGC GGC ACC GAA GGT TTA GGT TAC TTG TAT GCC GAG CGC ACC ATT
S V E I V D N G G T E G L G Y L Y A E R T I>

30200 30220 30240
GAC CCA AGT GAT TGG TTC TTC CAG TTC CAC TTC CAC CAA GAT CCG GTT ATG CCA GGC TCC TTA GGT
D P S D W F F Q F H F H Q D P V M P G S L G>

30260 30280 30300
GTT GAA GCA ATT ATT GAA ACC ATG CAA GCT TAC GCT ATT AGT AAA GAC TTG GGC GCA GAT TTC AAA
V E A I I E T M Q A Y A I S K D L G A D F K>

30320 30340 30360 30380
AAT CCT AAG TTT GGT CAG ATT TTA TCG AAC ATC AAG TGG AAG TAT CGC GGT CAA ATC AAT CCG CTG
N P K F G Q I L S N I K W K Y R G Q I N P L>

30400 30420 30440
AAC AAG CAG ATG TCT ATG GAT GTC AGC ATT ACT TCA ATC AAA GAT GAA GAC GGT AAG AAA GTC ATC
N K Q M S M D V S I T S I K D E D G K K V I>

30460 30480 30500
ACA GGT AAT GCC AGC TTG AGT AAA GAT GGT CTG CGC ATA TAC GAG GTC TTC GAT ATA GCT ATC AGC
T G N A S L S K D G L R I Y E V F D I A I S>

30520 30540 30560 30580
ATC GAA GAA TCT GTA T AAATCGGAGT GACTGTCTGG CTATTTTACT CAATTCTGT GTCAAAATG CTCACCTATA
I E E S V>

— exact match
w/ FAB A

Fig. 4
24/30

30600 30620 30640 30660
TTCATAGGCT GCGCGCTTTT TTCTGGAAAT TGAGCAAAAG TATCTGCGTC CTAACCTCGAT TTATAAGAAT GGTTTAATTG
30680 30700 30720 30740
AAAAGAACAA CAGCTAAGAG CCGCAAGCTC AATATAAATA ATTAAGGGTC TTACAAATA ATG AAT CCT ACA GCA ACT
M N P T A T>
30760 30780 30800
AAC GAA ATG CTT TCT CCG TGG CCA TGG GCT GTG ACA GAG TCA AAT ATC AGT TTT GAC GTG CAA GTG
N E M L S P W A V T E S N I S F D V Q V>
30820 30840 30860
ATG GAA CAA CAA CTT AAA GAT TTT AGC CGG GCA TGT TAC GTG GTC AAT CAT GCC GAC CAC GGC TTT
M E Q Q L K D F S R A C Y V V N H A D H G F>
30880 30900 30920 30940
GGT ATT GCG CAA ACT GCC GAT ATC GTG ACT GAA CAA GCG GCA AAC AGC ACA GAT TTA CCT GTT AGT
G I A Q T A D I V T E Q A A N S T D L P V S>
30960 30980 31000
GCT TTT ACT CCT GCA TTA GGT ACC GAA AGC CTA GGC GAC AAT AAT TTC CGC CGC GTT CAC GGC GTT
A F T P A L G T E S L G D N N F R R V H G V>
31020 31040 31060
AAA TAC GCT TAT TAC GCA GGC GCT ATG GCA AAC GGT ATT TCA TCT GAA GAG CTA GTG ATT GCC CTA
K Y A Y Y A G A M A N G I S S E E L V I A L>
31080 31100 31120 31140
GGT CAA GCT GGC ATT TTG TGT GGT TCG TTT GGA GCA GCC GGT CTT ATT CCA AGT CGC GTT GAA GCG
G Q A G I L C G S F G A A G L I P S R V E A>
31160 31180 31200
GCA ATT AAC CGT ATT CAA GCA GCG CTG CCA AAT GGC CCT TAT ATG TTT AAC CTT ATC CAT AGT CCT
A I N R I Q A A L P N G P Y M F N L I H S P>
31220 31240 31260
AGC GAG CCA GCA TTA GAG CGT GGC AGC GTA GAG CTA TTT TTA AAG CAT AAG GTA CGC ACC GTT GAA
S E P A L E R S V E L F L K H K V R T V E>
31280 31300 31320 31340
GCA TCA GCT TTC TTA GGT CTA ACA CCA CAA ATC GTC TAT TAC CGT GCA GCA GGA TTG AGC CGA GAC
A S A F L G T P Q I V Y Y R A A G L S R D>
31360 31380 31400
GCA CAA GGT AAA GTT GTG GTT GGT AAC AAG GTT ATC GCT AAA GTA AGT CGC ACC GAA GTG GCT GAA
A Q G K V V G N K V I A K V S R T E V A E>
31420 31440 31460
AAG TTT ATG ATG CCA GCG CCC GCA AAA ATG CTA CAA AAA CTA GTT GAT GAC GGT TCA ATT ACC GCT
K F M M P A P A K M L Q K L V D D G S I T A>
31480 31500 31520
GAG CAA ATG GAG CTG GCG CAA CTT GTA CCT ATG GCT GAC GAC ATC ACT GCA GAG GCC GAT TCA GGT
E Q M E L A Q L V P M A D D I T A E A D S G>
31540 31560 31580 31600
GGC CAT ACT GAT AAC CGT CCA TTA GTA ACA TTG CTG CCA ACC ATT TTA GCG CTG AAA GAA GAA ATT
G H T D N R P L V T L L P T I L A L K E E I>
31620 31640 31660
CAA GCT AAA TAC CAA TAC GAC ACT CCT ATT CGT GTC GGT TGT GGT GGC GGT GTG GGT ACG CCT GAT
Q A K Y Q Y D T P I R V G C G V G T P D>
31680 31700 31720
GCA GCG CTG GCA ACG TTT AAC ATG GGC GCG GCG TAT ATT GTT ACC GGC TCT ATC AAC CAA GCT TGT
A A L A T F N M G A A Y I V T G S I N Q A C>
31740 31760 31780 31800
GTT GAA GCG GGC GCA AGT GAT CAC ACT CGT AAA TTA CTT GCC ACC ACT GAA ATG GCC GAT GTG ACT
V E A G A S D H T R K L L A T T E M A D V T>

0.89

Fig. 4
25/30

31820 31840 31860
ATG GCA CCA GCT GCA GAT ATG TTC GAG ATG GGC GTA AAA CTG CAG GTG GTT AAG CGC GGC ACG CTA
M A P A A D M F E M G V K L Q V V K R G T L>
31880 31900 31920
TTC CCA ATG CGC GCT AAC AAG CTA TAT GAG ATC TAC ACG CGT TAC GAT TCA ATC GAA GCG ATC CCA
F P M R A N K L Y E I Y T R Y D S I E A I P>
31940 31960 31980 32000
TTA GAC GAG CGT GAA AAG CTT GAG AAA CAA GTA TTC CGC TCA AGC CTA GAT GAA ATA TGG GCA GGT
L D E R E K L E K Q V F R S S L D E I W A G>
32020 32040 32060
ACA GTG GCG CAC TTT AAC GAG CGC GAC CCT AAG CAA ATC GAA CGC GCA GAG GGT AAC CCT AAG CGT
T V A H F N E R D P K Q I E R A E G N P K R>
32080 32100 32120
AAA ATG GCA TTG ATT TTC CGT TGG TAC TTA GGT CTT TCT AGT CGC TGG TCA AAC TCA GGC GAA GTG
K M A L I F R W Y L G L S S R W S N S G E V>
32140 32160 32180
GGT CGT GAA ATG GAT TAT CAA ATT TGG GCT GGC CCT GCT CTC GGT GCA TTT AAC CAA TGG GCA AAA
G R E M D Y Q I W A G P A L G A F N Q W A K>
32200 32220 32240 32260
GGC AGT TAC TTA GAT AAC TAT CAA GAC CGA AAT GCC GTC GAT TTG GCA AAG CAC TTA ATG TAC GGC
G S Y L D N Y Q D R N A V D L A K H L M Y G>
32280 32300 32320
GCG GCT TAC TTA AAT CGT ATT AAC TCG CTA ACG GCT CAA GGC GTT AAA GTG CCA GCA CAG TTA CTT
A A Y L N R I N S L T A Q G V K V P A Q L L>
32340 32360 32380 32400
CGC TGG AAG CCA AAC CAA AGA ATG GCC TA ATACACTTAC AAAGCACCAG TCTAAAAAGC CACTAATCTT
R W K P N Q R M A>
32420 32440 32460 32480
GATTAGTGGC TTTTATTATT GTGGTCAATA TGAGGCTATT TAGCCTGTAA GCCTGAAAAT ATCAGCACTC TGACTTTACA
32500 32520 32540 32560
AGCAAATTAT AATTAAGGCA GGGCTCTACT CATTTATACT GCTAGCAAAC AAGCAAGTTG CCCAGTAAAA CAACAAGGTA
32580 32600 32620 32640
CCTGATTAT ATCGTCATAA AAGTTGGCTA GAGATTCGTT ATTGATCTTT ACTGATTAGA GTCGCTCTGT TTGAAAAAG
32660 32680 32700 32720
GTTTCTCGTT ATCATCAAAA TACACTCTCA AACCTTTAAT CAATTACAAC TTAGGCTTTC TGCGGGCATT TTTATCTTAT
32740 32760 32780 32800
TTGCCACAGC TGTATTGGCC TTTAGGTTTT GGGTGCAACT ACCATTAATT GAGGCCTCAT TAGTTAAATT ATCTGAGCAA
32820 32840 32860
GAGCTCACCT CTTTAAATTA CGCTTTTCAG CAA ATG AGA AAG CCA CTA CAA ACC ATT AAT TAC GAC TAT GCG
M R K P L Q T I N Y D Y A>
32880 32900 32920
GTG TGG GAC AGA ACC TAC AGC TAT ATG AAA TCA AAC TCA GCG AGC GCT AAA AGG TAC TAT GAA AAA
V W D R T Y S Y M K S N S A S A K R Y Y E K>
32940 32960 32980 33000
CAT GAG TAC CCA GAT GAT ACG TTC AAG AGT TTA AAA GTC GAC GCA GTA TTT ATA TTC AAC CGT ACA
H E Y P D D T F K S L K V D G V F I F N R T>
33020 33040 33060
AAT CAG CCA GTT TTT AGT AAA GGT TTT AAT CAT AGA AAT GAT ATA CCG CTG GTC TTT GAA TTA ACT
N Q P V F S K G F N H R N D I P L V F E L T>
33080 33100 33120
GAC TTT AAA CAA CAT CCA CAA AAC ATC GCA TTA TCT CCA CAA ACC AAA CAG GCA CAC CCA CCG GCA
D F K Q H P Q N I A L S P Q T K Q A H P P A>

Fig. 4
26/30

33140 33160 33180 33200
AGT AAG CCG TTA GAC TCC CCT GAT GAT GTG CCT TCT ACC CAT GGG GTT ATC GCC ACA CGA TAC GGT
S K P L D S P D D V P S T H G V I A T R Y G>
33220 33240 33260
CCA GCA ATT TAT AGC TCT ACC AGC ATT TTA AAA TCT GAT CGT AGC GGC TCC CAA CTT GGT TAT TTA
P A I Y S S T S I L K S D R S G S Q L G Y L>
33280 33300 33320
GTC TTC ATT AGG TTA ATT GAT GAA TGG TTC ATC GCT GAG CTA TCG CAA TAC ACT GCC GCA GGT GTT
V F I R L I D E W F I A E L S Q Y T A A G V>
33340 33360 33380 33400
GAA ATC GCT ATG GCT GAT GCC GCA GAC GCA CAA TTA GCG AGA TTA GGC GCA AAC ACT AAG CTT AAT
E I A M A D A A D A Q L A R L G A N T K L N>
33420 33440 33460
AAA GTA ACC GCT ACA TCC GAA CGG TTA ATA ACT AAT GTC GAT GGT AAG CCT CTG TTG AAG TTA GTG
K V T A T S E R L I T N V D G K P L L K L V>
33480 33500 33520
CTT TAC CAT ACC AAT AAC CAA CCG CCG CCG ATG CTA GAT TAC AGT ATA ATA ATT CTA TTA GTT GAG
L Y H T N N Q P P P M L D Y S I I I L L V E>
33540 33560 33580
ATG TCA TTT TTA CTG ATC CTC GCT TAT TTC CTT TAC TCC TAC TTC TTA GTC AGG CCA GTT AGA AAG
M S F L L I L A Y F L Y S Y F L V R P V R K>
33600 33620 33640 33660
CTG GCT TCA GAT ATT AAA AAA ATG GAT AAA AGT CGT GAA ATT AAA AAG CTA AGG TAT CAC TAC CCT
L A S D I K K M D K S R E I K K L R Y H Y P>
33680 33700 33720
ATT ACT GAG CTA GTC AAA GTT GCG ACT CAC TTC AAC GCC CTA ATG GGG ACG ATT CAG GAA CAA ACT
I T E L V K V A T H F N A L M G T I Q E Q T>
33740 33760 33780
AAA CAG CTT AAT GAA CAA GTT TTT ATT GAT AAA TTA ACC AAT ATT CCC AAT CGT CGC GCT TTT GAG
K Q L N E Q V F I D K L T N I P N R R A F E>
33800 33820 33840 33860
CAG CGA CTT GAA ACC TAT TGC CAA CTG CTA GCC CGG CAA CAA ATT GGC TTT ACT CTC ATC ATT GCC
Q R L E T Y C Q L L A R Q Q I G F T L I I A>
33880 33900 33920
GAT GTG GAT CAT TTT AAA GAG TAC AAC GAT ACT CTT GGC CAC CTT GCT GGG GAT GAA GCA TTA ATA
D V D H F K E Y N D T L G H L A G D E A L I>
33940 33960 33980
AAA GTG GCA CAA ACA CTA TCG CAA CAG TTT TAC CGT GCA GAA GAT ATT TGT GCC CGT TTT GGT GGT
K V A Q T L S Q Q F Y R A E D I C A R F G G>
34000 34020 34040 34060
GAA GAA TTT ATT ATG TTA TTT CGA GAC ATA CCT GAT GAG CCC TTG CAG AGA AAG CTC GAT GCG ATG
E E F I M L F R D I P D E P L Q R K L D A M>
34080 34100 34120
CTG CAC TCT TTT GCA GAG CTC AAC CTA CCT CAT CCA AAC TCA TCA ACC GCT AAT TAC GTT ACT GTG
L H S F A E L N L P H P N S S T A N Y V T V>
34140 34160 34180
AGC CTT GGG GTT TGC ACA GTT GTT GCT GTT GAT GAT TTT GAA TTT AAA AGT GAG TCG CAT ATT ATT
S L G V C T V V A V D D F E F K S E S H I I>
34200 34220 34240
GGC AGT CAG GCT GCA TTA ATC GCA GAT AAG GCG CTT TAT CAT GCT AAA GCC TGT GGT CGT AAC CAG
G S Q A A L I A D K A L Y H A K A C G R N Q>
34260 34280 34300 34320
TTG TCA AAA ACT ACT ATT ACT GTT GAT GAG ATT GAG CAA TTA GAA GCA AAT AAA ATC GGT CAT CAA

Fig. 4
27/30

L S K T T I T V D E I E Q L E A N K I G H Q>
34340 34360 34380 34400
GCC TAA ACTCGTTCGA GTACTTTCCC CTAAGTCAGA GCTATTTGCC ACTTCAAGAT GTGGGTACAA GGCTTACTCT
A>
34420 34440 34460 34480
TTCAAAACCT GCATCAATAG AACACAGCAA AATACAATAA TTTAAGTCAA TTTAGCCTAT TAAACAGAGT TAATGACAGC
34500 34520 34540 34560
TCATGGTGGC AACTTATTAG CTATTTCTAG CAATATAAAA ACTTATCCAT TAGTAGTAAC CAATAAAAAA ACTAATATAT
34580 34600 34620 34640
AAACTATTTT AATCATTATT TTACAGATGA TTAGCTACCA CCCACCTTAA GCTGGCTATA TTCGCACTAG TAAAAATAAA
34660 34680 34700 34720
CATTAGATCG GGTTACGATC AATTTACGAG TCTCGTATAA AATGTACAAT AATTCACTTA ATTTAATACT GCATATTTTT
34740 34760 34780 34800
ACAAGTAGAG AGCGGTGATG AAACAAAATA CGAAAGGCTT TACATTAATT GAATTAGTCA TCGTGATTAT TATTCTCGGT
34820 34840 34860 34880
ATACTTGCTG CTGTGGCACT GCCGAAATTC ATCAATGTTT AAGATGACGC TAGGATCTCT GCGATGAGCG GTCAGTTTTC
34900 34920 34940 34960
ATCATTTGAA AGTGCCGTAA AACTATACCA TAGCGGTTGG TTAGCCAAAG GCTACAACAC TCGCGTTGAA AAGCTCTCAG
34980 35000 35020 35040
GCTTTGGCCA AGGTAATGTT GCATCAAGTG ACACAGGTTT TCCGTACTCA ACATCAGGCA CGAGTACTGA TGTGCATAAA
35060 35080 35100 35120
GCTTGTGGTG AACTATGGCA TGGCATTACC GATACAGACT TCACAATTGG TGGCGTTAGT GATGGCGATC TAATGACTGC
35140 35160 35180 35200
AGATGTCGAT ATTGCTTACA CCTATCGTGG TGATATGTGT ATCTATCGCG ATCTGTATTT TATTCAGCGC TCATTACCTA
35220 35240 35260 35280
CTAAGGTGAT GAACATACAA TTTAAACTG GTGAAATAGA AATTATTGAT GCTTCTTACA ACCCTGACGG CTCAACTGGT
35300 35320 35340 35360
CAATTACCAT AAATTGGCG CTTATCTAAG TTGTAAGTGC TCTGACCGAC ACAAATAATG TCGTTTCTCA GCATATATCA
35380 35400 35420 35440
AAATACACAG CAAAAATTG GGGTTAGCTA TATAGCTAAC CCCAAATCAT ATCTAACTTT AACTGCATC TAATTCCAAA
35460 35480 35500 35520
CAGTATCCAG CCAAAAGCCT AAACATTGT TGACTCAGCG CTAAATATG CGATGCAACA AACAAGTCTT GGATCGCAAT
35540 35560 35580 35600
ACCTGAGCTA TCAAAAATGG TCACCTCATC AGCACTTTGA CGTCCTGTTG CGGACTCGTT TATCACTGA CCAATCTCAA
35620 35640 35660 35680
TTATCGGCGT ATTTCTGCTA TGTGAAACT CACCAATAAC AATAGATTGA GAAGCAAAGT CGCAAAACAA GCGAGCATGA
35700 35720 35740 35760
CTATATAGGT CAGTTGGCAA CTCTTGCTTA CCCACTTTAT CAGCGCCCAT TGCAGAAATA TCGGTTCTTG CTGTATCCCA
35780 35800 35820 35840
CTGCGCTTCA AATAAAGGCG CTTGAGCTGT GGTGCTGTG ATAATAATAT CTGCTTGTTC ACAAGCAGCT TGTGCATCAC
35860 35880 35900 35920
AAGCTTCGGC ATTAATGCCT TTTTCTAATA AACGCTTAAC CAAGTTTTC GTTTGTCTAG CACTACGGCC AACTACCAAT
35940 35960 35980 36000
ACCTTAGTTA ATGAACGAAC CTTGCTCACT GCTAGCACTT CATATTCAGC CTGATGACCG GTACCAAAAA CAGTTAATAC
36020 36040 36060 36080

Fig. 4
28/30

CGTAGCATCT TCTCTCGCGA GGTAACAC TGCTACTGCA TCGGCAGCAC CAGTGC GGTA AGCATTAAACG GTAGTGGCAG
36100 36120 36140 36160
CAATCACCGN CTGCAACATA CCGGTAAATG GATCGAGTAA AAATACGTTA GTGCCGTGGC ATGGTAAACC ATGTTTATGG
36180 36200 36220 36240
TTATCAGGCC AATAGCTGCC TGTTCCTCAG CCGACAAGGT TTGGCGTTGA AGCCGACTTT AATGAGAACA TTTCATTAAAG
36260 36280 36300 36320
GTTCGCGCCC TGTGCATTAA CTACCGGGAA CAAGGTTGCT TTATCATCTA CGGCAGCGAC AAACGCTTCT TTAACAGCGA
36340 36360 36380 36400
TATAAGCCAG CTCATGGGAG ATGAGCTTTG ATGTTTGGCG TTCAGTTAAA TAGATCATAT TACCACCCCT GCACTCGATT
36420 36440 36460 36480
CCAGATCTCA TAGCCACCAT TATCACCATC AGTATCAAAT ACATGGTACT GAGCGTGCAT TGAAGCTGTT GCACAGGCGT
36500 36520 36540 36560
GGTTCGGCAA AATATGTAGA CGACTACCTA CCGGGAAGCT CGCTAAATCA ATAACGCCGC CATCAACTGC TTCAATAATG
36580 36600 36620 36640
CCGTGCTCTT GATTAAACAGT TATAACCTGT AGACCTGATA ACACGTGACC GCTGTCGTCA CACACTAAAC CATAACCACA
36660 36680 36700 36720
ATCTTTTGGC TGCTCTGCAG TACCTCTATC ACCCGAAAGA GCCATCCAAC CCGCATCAAT GAAAATCCAG TTTTATCAG
36740 36760 36780 36800
GATTATGACC AATAACACTG GTCACTACCG TTGCGGCAAT ATCAGTTAAC TGACACACGT TTAGCCCTGC CATGACTAAA
36820 36840 36860 36880
TCGAAGAAGG TGTACACACC CGCTCTAACC TCGGTGATCC CATCAAGGTT TTGATAGCTT TGCCTGTTG GTGTTGAACC
36900 36920 36940 36960
AATACTAACG ATGTCACATT GCATACCCGC TCGCGGAATG CGTCAGCAGC TTGTACAGCC GCTGCAACTT CATTTTGGCG
36980 37000 37020 37040
CGCATCAATT AATTGCTGTT TTTCAAAACA TTGATATGAC TCACCAGCGT GAGTNAGTAC GCCGTGAAAA CTCGCTGCGC
37060 37080 37100 37120
CAGACGTTAG TATCTGAGCA ATTTCAATCA ACTTATCGGC TTCCGGTGA ATACCACCAC GATGGCCATC ACAATCAATT
37140 37160 37180 37200
TCAATTAATG CTGSTATTTG GCAGTCATAA GAACCACAGA AATGATTTAG CTGATGCGCT TGCTCAACAC TATCAAGTAA
37220 37240 37260 37280
AACTCTTGCA TTAATACCTT GGTCACACAT TTTAGCAATA CGCGGCAACT TACCATCGGC AATACCTACT GCATAAATAA
37300 37320 37340 37360
TGTCTGTGTA ACCTTTAGAT GCTAAGGCCT CGGCCTCTTT TACCGTTGAT ACAGTGACTG CTGAGTTTTT AGTGGGTAAT
37380 37400 37420 37440
AAAAACTCGG CTGCTTCAAG TGATCTTAAC GTTTTAAAT GCGGTCTTAG GTTTGCACCT AATCCTTCAA TTTTGGCG
37460 37480 37500 37520
TAGTTGACTG AGGTTATTAA TAAATACTGG CTTATTTACA TATAAAAACG GTGTATCAAT TGCTTGATAC TGACTTTGCT
37540 37560 37580 37600
GAGTCGTGGA AAGTATTTGA GTAGATGGCA TCTTTAATAT CCTAGTTCAT CAATCAATCT AACAAGTTTG ATGCCTAGCC
37620 37640 37660 37680
ACAGTGGCTT GTATTCATGA TGCTTTGGAA AATGCTTATA TTCAAAGTAT TTGAAAGACA TCAAACCTCT TGTTTAATGC
37700 37720 37740 37760
TCAGTATCCA CCAGCAGCGA TTTATTTTAT ATTAACATAT ATCAAGATAT AGATTAGGTT CAAACCAAAT GATTAGTACT
37780 37800 37820 37840
GAAGATCTAC GTTTTATCAG CGTAATCGCC AGTCATCGCA CCTTAGCTGA TGCCGCTAGA AACTAAATA TCACGCCACC

Fig. 4
29/30

37860

37880

ATCAGTGACA TTAAGGTTGC AGCATATTGA AAAGAAACTA TCGATTAGCC TGATC

Fig. 4
30/30

10 20 30 40 50 60
AATAGATCGACTCGCAAAAGTTGCTTAAGATAGTGTCAATATAGCTTCTTATTTGTAAAT

70 80 90 100 110 120
ATTGTTTTTTATGTGTAAACATGTTTAGTGTGTGTAAATGCTGTTAATTATCCTTTTGGG

130 140 150 160 170 180
ATTGTAATAGCTGATGTTGCTGGCTAATGAGTACTTTTAGTTCGGCAATATCTTGCTTTA

190 200 210 220 230 240
AATCGCTAACTTCAGTTTTTAATTCACCCACACTTGTTGTATTTTTAAGGCTCTCTTCCC

250 260 270 280 290 300
CACCATCGACAAACCAGGATGATATGAAACCGTAAACGTACCAAAGAGACCGACACCTG

310 320 330 340 350 360
CAGTCATGAGTAATGCCGCAATGATACGTCCGCCAGTGGTGACGGGGTAGTAGTCACCGT

370 380 390 400 410 420
AACCAACAGTCGTTATTGTACAAATGACCACCAAAGTGCCTCGATGCCGTTATTGATGT

430 440 450 460 470 480
TACTGCCTACTTGATCCTGTTCTAACAATAAAATACCGATAGCACCAAAGGTGACAAGGA

490 500 510 520 530 540
TGAAGGATATCGCAGATACCAGCGAAAAGGTGGCTTTAAACCGATGTTCAAAAATCATTT

550 560 570 580 590 600
TTAAGATAATTTTTGATGAGCGTATATTCTGAATAGATCTTAATACTCTAGCGATACGAA

610 620 630 640 650 660
TTATGCGAATAAACTGCAGTTGCTCGACCATCGGAATACTCGACAGTAGGTCAATCCAAC

670 680 690 700 710 720
CCCATTTTCATAAACTGAAATTTATTCTCAGCTTGGTGAAAGCGAATTACAAAGTCAGTGA

730 740 750 760 770 780
AAAAGAATAAGCAAATCGTATTATCTACGCTCGTTAATATTTTCAGTGACGTTACTTGAAA

790 800 810 820 830 840
AGGTAAAAATAAGTTGCAGTAGTGATGATACGACCACATGAAGTGATAAAATAAGCATGA

850 860 870 880 890 900
AAATCTGAAATGGATTTACATCACTGTTGTTTTTGGTGCCACTTTTAAGGTTCTGTTTTCA

910 920 930 940 950 960
CAATCTGCTGCCCTCGGTTTCATTGATTTTTGTTAATATAAACCTTAGTCAGTAGCAAGACAA

970 980 990 1000 1010 1020
AATATATTTACATCAATGTCATCGTATTATTCAACCGCGCGTCGTGTATTCAGACCAAGA

1030 1040 1050 1060 1070 1080
TCGTTGTATATGTTAGTCATGTAGCGATGAGATTATCATGCGACAGGAGAGAATTATGTT

1090 1100 1110 1120 1130 1140
TGTTATTATTTTTTACGTACCTAAAGTTAATGTTGAAGAAGTAAACAGGCGTTATTTAA

Fig. 5

1150 1160 1170 1180 1190 1200
CGTCGGAGCTGGCACCATCGGTGATTATGATAGTTGTGCTTGGCAATGTTTGGGGACTGG

1210 1220 1230 1240 1250 1260
GCAGTTCCAACCTTTACTTGGTAGCCAGCCACATATTGGTAAGCTAAATGAGGTTGAATT

1270 1280 1290 1300 1310 1320
CGTTGATGAGTTTAGAGTAGAAATGGTTTGTGCGAGCAGAAAATGTAAGGGCAGCAATAAA

1330 1340 1350 1360 1370 1380
TGCACTTATTGCTGCGCACCCTTATGAAGAACCTGCTTATCATATTCTGCAAACATTGAA

1390 1400 1410 1420 1430 1440
TCTTGATGAGTTACCTTAAGTTAGATGCACTGCACTTAATTGGTTCGCTGTGCTAGGTTA

1450 1460 1470 1480 1490 1500
GCAATTAGCAATTTTGACCATGTTAGCGATAGTTTGGCACAAGTGATCGATATTAAACT

1510 1520 1530 1540 1550 1560
ATCCGATTTCAGATCCCATTTTTACTGCTGAATTAGGTTTCATTACACTTGTTCTAGTGGT

1570 1580 1590 1600 1610 1620
TTTTCCCGACAGGTGTAACCTCTGTTACTTGCGTAAGGTTGATAATCTCTACCGCATTGGC

1630 1640 1650 1660 1670 1680
AGGAGTTACACCTGCACCAGGCATAATACTAATTCTACCATCTGCTTGGTTAACTAACGT

1690 1700 1710 1720 1730 1740
TTGGATTAAGGCGCAGCCTTCTAGCGCTTGAGCTTGTTGACCAGAGGTTAAAATACGCTC

1750 1760 1770 1780 1790 1800
ACAACCAGCAGTGATCAAGGTCTCCAAGGCTTGTTGTGGATCATTACACAAGTCGAAAGC

1810 1820 1830 1840 1850 1860
GCGGTGGAAGGTTACGCCGAGATCACGTGATGCCACCATTAAAGCGTTTTAAAGCTGGCTC

1870 1880 1890 1900 1910 1920
GTCAATATTACCATCTGCTGTAAACGCGCCAATAACGACCCCTTGGACACCGAGTAACCTT

1930 1940 1950 1960 1970 1980
CATGAATTTGATGTGCGAAACCATAATATCAACTTCTTGTTTCGCTATATACAAAATCACC

1990 2000 2010 2020 2030 2040
GGCGCGAGGGCGAATAATGGCATAAATGGGGATCGTTGCTAGATCAATAGACTTTTGTAC

2050 2060 2070 2080 2090 2100
AAAACCTGCGTTGGCGGTCAAGCCACCTAATGCTAATGCCGAGCACAACCTCAATACGATC

2110 2120 2130 2140 2150 2160
GGCGCCAGATGCTTGAGCCGTCAGCAGTGATTCTATATTATCGACACATACTTCTATTGT

2170 2180 2190 2200 2210 2220
CATTGTCAATACTTCTCTTTAAAAAGTTTATTAAAAATAATAAGCCAGCATAAGTCGT

2230 2240 2250 2260 2270 2280
TTTATACAATATGAAAGGGGAAAAGGCGACTTAGCTCGCCTAGATCAATTATTATGGCAG

Fig. 5

2290 2300 2310 2320 2330 2340
AATACTGCCGTATTGTGATTAGAAAGACAGTTTTTTAAGCTCAATAGCCGTTATCGCGTT

2350 2360 2370 2380 2390 2400
GTTATCTACCATCGTGTAACTTTTCTGGCCTGGGTGCTTTATTAACTGTTTCAGTGGC

2410 2420 2430 2440 2450 2460
TGGATTAGGGTGAAATGATTCTTTTTTCAAATCTGTTTTTTTGATTGAACGTACCTGT

2470 2480 2490 2500 2510 2520
AATGTCTTGCTGCTCACGAAGACGTACAAATATTGGTTGCGCATAGCTTGGTAGTGCCGC

2530 2540 2550 2560 2570 2580
ATTGACATGTTGATAGAATTCAGACGCTGAAAATTCATGAATAGGGCAATTCAAAGTCAG

2590 2600 2610 2620 2630 2640
CGCGACCATGCGCTGCTCGGCCATCGTGATGTGGGAGCTTGACACCATAAGCCACACTTTG

2650 2660 2670 2680 2690 2700
CTCAATTTGCACAAAATCGTTAACTTGAGCTTCTACTTGCGTCGTGGCGACATTTTCACC

2710 2720 2730 2740 2750 2760
TTTCCAGCGGAATGTATCACCTAATCTATCCACAAAGGAAATATGGCGATAACCTTGGA

2770 2780 2790 2800 2810 2820
ATGAACGAGATCGCCGGTATTAAAATAACAGTCACCGTCTTTTAATACTGACTTAAATAG

2830 2840 2850 2860 2870 2880
CTTTTTATTACTTTTCGTTGTCATCGGTATAACCATCAAATGGTGAACGTTTAGTTATCTT

2890 2900 2910 2920 2930 2940
TGTTAGCAGTAGCCCTGTTTCTCCCGTTTTTACTTTGGTCATTTTCCCTTTCGCATTATA

2950 2960 2970 2980 2990 3000
CACAGGTTTGTCAATTGTCAATATCATATTGTATGACGGTAAAAGCAAGTGGAGTAACCCC

3010 3020 3030 3040 3050 3060
CGCTGTATGCGGTAAGTTCAGCGCATTGGGAGAACACAAGATTACACTCACTGGCGCCATA

3070 3080 3090 3100 3110 3120
GAATTCATTAATATGCTCGATCCCAAAACGTTGTTGGAAATGATCCCAAATTTTCGGGGCG

3130 3140 3150 3160 3170 3180
TAATCCATTACCTATGATTTTCTTTATATTATGCTGTTTGTCTTTATTGCTAGGCGGTAC

3190 3200 3210 3220 3230 3240
ATTTAATAAATAACGGCAGAGCTCGCCGATGTAAGTAAACGCAGTGGCATTATGAGCACG

3250 3260 3270 3280 3290 3300
AACTTCATCCCAAAAGCGACTTGAAGTGAATTTTTCAGAAAGTGCGAGGGTTGCTGCGCT

3310 3320 3330 3340 3350 3360
ACCAAACACGGCGCTTAATGACACTGTGAGTGCATTGTTATGGTATAGGGGGAGTGATAA

3370 3380 3390 3400 3410 3420
ATACAATACATCATCAGCTGTTAAGCGTAATGATGCCATCCCCATGCCTGCCATGGATTT

Fig. 5

3430 3440 3450 3460 3470 3480
AAACCAACGGTGATGGCTCATTCTTGCTGCTTTTGGCAGTCCAGTTTTTCCCGAGGTAAA

3490 3500 3510 3520 3530 3540
GATATAAAACGCGCAATGCTTAAGCTGTATTTGTGCTGTTGATTGAGGGTTCAATACTGA

3550 3560 3570 3580 3590 3600
ATATCCTGCGACTAGTGTAGATATGTTTTATAACCATCACTCATGTCTGGCGTTTCTAA

3610 3620 3630 3640 3650 3660
AGCGGGTACGTAAAAGACATTCTGTTGTAATGTCGATGACAAATTGGTTTCAATATTATT

3670 3680 3690 3700 3710 3720
AATGGCGGATGTGTATAGTTCATCTGCGATGAGTAATTTGGTATCGACCACGCTAAGACT

3730 3740 3750 3760 3770 3780
ATGTTTCGAGGATTGAATCCCGTTGTGTCGTATTTATCATAACAAGCAATCGCGCCAAGCTT

3790 3800 3810 3820 3830 3840
GACAACTGCGAGGGCAATAATGATGGTTTCAGGCCTGTTATCGAGCATGATGGCGACTTT

3850 3860 3870 3880 3890 3900
ATCATTTTTACCAATGCCGTATTCATGAAGGAAATGGGCATATTGATTTGCTTGCTTATT

3910 3920 3930 3940 3950 3960
CAATGAATCGTAACATAACGCTGGTCTTTAAATTGTATTGCGATCAAGTCAGAGTTATT

3970 3980 3990 4000 4010 4020
GACAGCTTGCTGCTCTAGTAATAAACCAATAGACATAAAACGTTCCGGGCTTTGCTTGTTG

4030 4040 4050 4060 4070 4080
TAAGTGCCATAAGCCTTTGATGATTGGCTTTGGGGTTTTTAATAGATTGATGGTACTTTT

4090 4100 4110 4120 4130 4140
CAGGAATTGTTTGCCGGTTATAACAGTCATAAGCTAATTCTTTTATCAAGAAGAGGGGT

4150 4160 4170 4180 4190 4200
TATGACACCAAATAAATGGGTCACGCGTTGGTTTAATTTGGTTAGACTAAATGTGTGTT

4210 4220 4230 4240 4250 4260
TTGCTGTGATAATGCGACGTTCAAACAACTTGAGAAGGTAAAAAATAGCATTTTTTAAA

4270 4280 4290 4300 4310 4320
TTGAACATCAATACTAATGTGTTGAATATCAATCAAGTTTTCTAACTGTGCGAGCACGCG

4330 4340 4350 4360 4370 4380
TGCTTTAGCAAACATGCCATGTGCTATTGCTGTTTTAAACCCCATAGTTTCGCTGGGAT

4390 4400 4410 4420 4430 4440
AAAATGTAAATGGATTGGATTGTGTCTTTGGAGATATAAGCATATTTATATACGTCAA

4450 4460 4470 4480 4490 4500
AGGACTAAATTTAAACAATGAAATCGGCTCGTAAGCATAATTCGCTGGCGTATTTACTAT

4510 4520 4530 4540 4550 4560
TTTCTCACCGCTGGAACGTTGAGATCGTTGGCACGTTTTTCGCTGTTTCGTTTTCTGTAA

Fig. 5

4570 4580 4590 4600 4610 4620
GAATGTCGATGTACACTCCCACGCAAATTGTCCATCTACAAACACATCAATATGAGTATC

4630 4640 4650 4660 4670 4680
AATGAAACGTCCTGTATCCGTTATGTACTCCTTAATTACACGACATGTGCTCGTCAATAT

4690 4700 4710 4720 4730 4740
CGCGTTTAAATGCTATCGGTTGATGTTGTGTTATGCGATTTTCGATAATGGACTAGTCCTAA

4750 4760 4770 4780 4790 4800
TATAGATATCGGAAATTGTGTTGATGTCATGAGTTTCATCAATAATGGAAAGATCATCAC

4810 4820 4830 4840 4850 4860
AAATGGATAAGTAACCGGTACATAGTTTGTGTTATTAAACCCACAGCATTTAATATATTG

4870 4880 4890 4900 4910 4920
CTTTAAATTTTCGCTGATCTATTTTTGTCCACTGATACTAAATTGCTCAGTACACACTTG

4930 4940 4950 4960 4970 4980
TGTCGACCAAGTGTTTCATCAGTGTTTTAACAATTGTATTGACCACTGCTTTCACATATAA

4990 5000 5010 5020 5030 5040
AAGCGAGATAATCGGTTGCTTTGTTAACAGTGTGATCTGGTTAGCGTGCATTGAAATAAT

5050 5060 5070 5080 5090 5100
TCATATAAGAGTATGTAGCATTTATGTTAATATTTGTTTTGGAAGTTGAATTGGCGAAT

5110 5120 5130 5140 5150 5160
CCGTAATCGGTTTATGGCAGTTCGGTCAAATACTTCAGGTAAACTCGTTACTCATACCAT

5170 5180 5190 5200 5210 5220
TGATAGTGTTAAAGTGATTGACTGAATAAAGAATAGAGCTAAAAGTGGAAAAATTATGCA

5230 5240 5250 5260 5270 5280
AGATGCGGGTATGTTATTACGCATTGCTTATGAGGCAATGAAAGAGTTAGAGGTTGATGT

5290 5300 5310 5320 5330 5340
CATTGAAGTACTTTCTCGTTGTAACATAAGTGAAGAAGTACTGAATGATAAGGATCTTCG

5350 5360 5370 5380 5390 5400
CACACCTAATCATGCACAAACACATTTTTGGCAAGTATTAGAAGACATATCACAAGATCC

5410 5420 5430 5440 5450 5460
TAACATCGGCATTTCACTTGGTGAGAGAATGCCAGTGTTACGGGGCAGGTATTACAGTA

5470 5480 5490 5500 5510 5520
TCTTTTTCTCAGTAGTCCTACATTTGGTACTGGCTGGGAACGCGCAACAAAATACTTTTCG

5530 5540 5550 5560 5570 5580
ATTAATCAGTGATGCGGCGAGTGTCTTCTATCAAGATGGAAGGCTGTGAAGCGCGATTATC

5590 5600 5610 5620 5630 5640
TGTGAACCTTAGATGGTTTAGCGGAAGATGCGAATCGTCATTTGAATGATTGCCTAGTGAT

5650 5660 5670 5680 5690 5700
CGGTGCATTTAAATTTTGTATATGTGACAAAGGCGAATTTAAAGTAAGCAAAATAGC

Fig. 5

5710 5720 5730 5740 5750 5760
CTTTGCTCATGCTCGCCCGAAAGATATTACTGCCTATACCAATGTATTTACATGTCCGAT

5770 5780 5790 5800 5810 5820
TGAGTTTGCTGCCGAAGATAATTATATTTATTTTCGATGCTGATTACTCGAACGTCCTTC

5830 5840 5850 5860 5870 5880
TTCGCATGCGGAGCCTGAGCTATTTCGCCTTACACGATCAGCTTGCAAGCCGTAAAATAGC

5890 5900 5910 5920 5930 5940
CAAGTTAGAACTGCAAGATTTAGTGGATAAAGTACGTAAGGTTATTGCACAACAACCTGA

5950 5960 5970 5980 5990 6000
GTCTGGTGTGGTGACTTTAGAAAGTATCGCCACTGAACTTGACATGAAACCACGTATGCT

6010 6020 6030 6040 6050 6060
AAGAGCGAAGTTAGCTGACATTGATTATAACTTTAATCAAATACTCGCTGATTTTCGTTG

6070 6080 6090 6100 6110 6120
CGAGTTATCAAAAAAAGTGTGGCGAATACGGACGAGTCTATTGATCAGATTGTCTATCT

6130 6140 6150 6160 6170 6180
CACTGGTTTTTCTGAACCAAGTACTTTTTATCGTGCCTTTAAGCGCTGGGTTAAAATGAC

6190 6200 6210 6220 6230 6240
GCCAATTGAATATCGCCGTAGCAAACCTCGCGGTTAGGCATGCTAATCAACACGAGTCCTA

6250 6260 6270 6280 6290 6300
AAAATTCGCTGCTTAGTGATAGTGATAGTGATAGTGCTAGTAAGCCAAGTACAAAGC

6310 6320 6330 6340 6350 6360
GTTAAAGTTAAGTACTTGAGCGAACCATCAGACACCACTTACTAGATTAAAGCACCTATTA

6370 6380 6390 6400 6410 6420
ATGATTGACCACAAATTCTGATCGTATTGCCTGTGATCCCTGCAGCTTGAGGTTGCGCAA

6430 6440 6450 6460 6470 6480
AAAAAGCTATCGCTTCAGCAACATCAACTGGCTTACCACCTTGTTTTAATGAATTCATAC

6490 6500 6510 6520 6530 6540
GACGACCAGCTTCACGAACTGTAAATGGAATCGCTGCTGTCATTTTTGTTTCAATAAAGC

6550 6560 6570 6580 6590 6600
CTGGTGCAACAGCATTAAATGGTGATGTATTTGTCTGCAAGCGGAGTTTGCATTGCATCAA

6610 6620 6630 6640 6650 6660
CATAACCAATGACTGCGGCCTTAGACGTTGCATAATTAGTCTGACCAAAGTTACCCGCAA

6670 6680 6690 6700 6710 6720
TCCCACTCATCGAAGACACACAAACAATGCGGCCATAGTCGTTGAGCAGATCATCATTTA

6730 6740 6750 6760 6770 6780
GCAGTCGCTCATTGATTCTTTCCATTGCCGACAAGTTAATATCCATCAGTACATCCCAAT

6790 6800 6810 6820 6830 6840
GGTTATCCGGCATACGTGCTAGCGTTTTGTCTTTTGTGTACCCCGGCATTATGGACGATGA

Fig. 5

6850 6860 6870 6880 6890 6900
TATCAAGCGACTGTTCTCGCACAAAGTCAGCAATGATATTTGGGGCGTCAGCAGCGGTAA

6910 6920 6930 6940 6950 6960
TATCAGCAACAATGCTGCTACCTTTCAAGCAATGAGCTACTTTTCAAGGTCCTGTTTTA

6970 6980 6990 7000 7010 7020
ATGCCGGAATGTCTAAGCAAATAACATGTGCGCCATCACGGGCGAGTGTTTCAGCAATAG

7030 7040 7050 7060 7070 7080
CAGCCCCGATGCCACGTGATGCACCAGTGACAAGTGCTGTCTTTCCTTGTAATGGTTTTG

7090 7100 7110 7120 7130 7140
CCGTGTTACTTGTTCGTTAATAACTTCGTTAATAACTTCGTTAATAACTTCGTTAATAG

7150 7160 7170 7180 7190 7200
CCCCATTAATCGAACC GG GTTTTACGTTAATAACCTGTGCTGAGATATAGGCTGATTTTG

7210 7220 7230 7240 7250 7260
CTGAGGTTAAGAAACGTAGCGGGGCCTCTAATAATTGCTCACTACCAGGTTGTACATAGA

7270 7280 7290 7300 7310 7320
TAAGTTGACAGGTACTACCATTCTTGCCTATTTCTTTGGCGACACTGCGACAAAACCTT

7330 7340 7350 7360 7370 7380
CTAAAGATCTTTGTACAGTCGCGTAGCTTACATCGTCAAGATGTTCACTCGGATGACCTA

7390 7400 7410 7420 7430 7440
ACACGATCACTCTGCTGCATGGCGAGAGCTGCTTAATTACAGGTTGAAAAAACGATGTA

7450 7460 7470 7480 7490 7500
ATGCACTTAATTGCTTGCTGTTCTTAATGCCTGAGGCGTCGAAGATAATACCGTTGAAGC

7510 7520 7530 7540 7550 7560
GATCTGTTTTAGCGATAGCATTAAAGGCTAATAGGTGTCGCGACTAAAGACGTTTGATTAA

7570 7580 7590 7600 7610 7620
ATTCAATATTAAGATCGGCTAACGCTGACGTGTTATTAGGATAAGAAATCGTGACTTCAG

7630 7640 7650 7660 7670 7680
CATCTTTAAATGTGTTAAGAATGGGTTTAATTAATTTGCTGTTGCTGGCTGCGCCGATGA

7690 7700 7710 7720 7730 7740
GTAAGTTGCCAGAGATGAGATCGGTTCCCTGATCGTAGCGTGTTAACGTAACCGGTCGTG

7750 7760 7770 7780 7790 7800
GCAGATTAAGCGCTTTAAATAAACCTGATGTCCACTTGCCATTAGCGAGTTTTGCGTATG

7810 7820 7830 7840 7850 7860
TATCCGTCATTTTCTAATCCTTGTTATAGTGAACAGTTTGAATCTCGAAGATGTACATGT

7870 7880 7890 7900 7910 7920
GTTAAAAATTATCTGATAGCTATGACTTATCTGCCACTACGTAATAATAAATAGACCAGT

7930 7940 7950 7960 7970 7980
TCATTACATCGTTAATCGATATAGTATAACTAAATACTAAGTAAATTATAATGATAAGAC

Fig. 5

M

7990 8000 8010 8020 8030 8040
TGTTATCGTACTCGGATCAAACCTCTGATCAGCAAATAATCAAATTAGAGTTTTTATTTTA

8050 8060 8070 8080 8090 8100
AACTTGATCAACAATGTTACATTAATGTATCTTACGTCTAATGTGCTACGGGCATATTT

8110 8120 8130 8140 8150 8160
AAGTCACTAAATTAAAGGAATAAACCATGACAGGTCAAACAATAAGAAGAGTAGCAATTA

8170 8180 8190 8200 8210 8220
TCGGCGGTAACCGTATCCCGTTTGACAGTTCAAATACAGCGTATTCAAACCTAAGTAACC

8230 8240 8250 8260 8270 8280
AAGATATGCTGACGGAACTATCCGTGGCTTGGTGGTTAAATATAACCTACGTGGTGAAC

8290 8300 8310 8320 8330 8340
AACTGGGGGAAGTTGTTGCTGGTGCGGTAATTAAGCATTCTCGTGATTTTAACCTAACAC

8350 8360 8370 8380 8390 8400
GTGAAGCCGTGCTAAGTGCAGGTCTTGACACCTGAAACGCCTTGTTATGACATTCAACAAG

8410 8420 8430 8440 8450 8460
CTTGTTGGTACTGGTCTAGCTGCAGCTATCCAAGTAGCAAACAAAATTGCGCTTGGTCAA

8470 8480 8490 8500 8510 8520
TAGAAGCGGGTATTGCTGGTGGTTCTGATACGACATCAGATGCACCGATTGCAGTCAGTG

8530 8540 8550 8560 8570 8580
AAGGCATGCGTAGTGTATTACTTGAGCTTAATCGAGCTAAAACGGGTAAGCAACGTTTGA

8590 8600 8610 8620 8630 8640
AAGCACTATCTCGTCTACGTCTAAAACACTTTGCGCCACTAACGCCTGCAAATAAAGAGC

8650 8660 8670 8680 8690 8700
CGCGTACCAAAATGGCGATGGGCGATCATTGTCAAGTAACAGCGAAAGAGTGGAATATCT

8710 8720 8730 8740 8750 8760
CACGTGAAGCACAAGATGCATTGGCCTGCGCAAGTCATCAAAAATTAGCTGCAGCATATG

8770 8780 8790 8800 8810 8820
AAGAAGGTTTCTTTGATACGTTAGTTTACCTATGGCCGGCTTAACGAAAGATAACGTAT

8830 8840 8850 8860 8870 8880
TACGCGCAGATACAACAGTTGAGAACTGGCTAAATTGAAACCTTGTTTTGATAAAGTAA

8890 8900 8910 8920 8930 8940
ACGGCACTATGACGGCGGGTAACAGTACTAACCTTACCGATGGAGCATCAGCTGTATTAC

8950 8960 8970 8980 8990 9000
TTGCAAGTGAAGAAATGGGCAGCGGCACATAACTTACCAGTACAAGCTTATCTAACATTTG

9010 9020 9030 9040 9050 9060
GTGAAACGGCCGCTATCGACTTCGTTGATAAGAAAGAAGGTCTGTTAATGGCGCCTGCAT

9070 9080 9090 9100 9110 9120
ACGCAGTGCCAAAAATGTTGAAGCGTGCTGGCCTTACATTACAAGACTTCGATTACTATG

Fig. 5

9130 9140 9150 9160 9170 9180
AAATACATGAAGCATTGCTGCGCAGTTATTAGCAACGCTAGCAGCTTGGGAAGACGAAA

9190 9200 9210 9220 9230 9240
AATTCTGTAAAGAAAACTGGGTCTAGATGCTGCGCTTGGTTCAATTGATATGACCAAGT

9250 9260 9270 9280 9290 9300
TAAACGTGAAAGGGAGTAGCTTAGCCACGGGTCACCCATTGCGCGAACTGGTGGTCGTG

9310 9320 9330 9340 9350 9360
TTGTGCTACGCTAGCGCAATTACTTGATCAGAAAGGTTCAAGTCGTGGTTTGATCTCGA

9370 9380 9390 9400 9410 9420
TTTGTGCTGCTGGTGGTCAAGGTATCACGGCAATTTTAGAGAAATAAACGCACTGTTTAT

9430 9440 9450 9460 9470 9480
TATCTATTGATTAAGCTGTCCTGAGATACTGGATATTTTTAAATAAAACGCCAATACTGC

9490 9500 9510 9520 9530 9540
AGAGTATTGGCGTTTTTTTTGTAATACCAATTCTATATAACGGTGCATTTTAAACACTTA

9550 9560 9570 9580 9590 9600
ATTTCCGGCATTGGTATCATAAAAAGCAGCACCGAAGTGCTGCTTGATTGTAGATTAAAC

9610 9620 9630 9640 9650 9660
CTATTAAATAGAGAGGCTAGAATTAGTCTTCGTATGCTTCATTATGTACGCCAGCTGCA

9670 9680 9690 9700 9710 9720
CGACCCGATGGATCAGCATTGTTTTGGAACTTTTCATCCCAAGCTAATGCTTCTACAGTT

9730 9740 9750 9760 9770 9780
GAACAAGCAACGGATTACCAAACGGTACGCATTTGCTGCTGAATCACCTGGGAAGTGA

9790 9800 9810 9820 9830 9840
TCTTCAAAGATGGCAGATAGTAGTAACCTTCTTTCGTATCTGGTGTGTTAATTGGGAAC

9850 9860 9870 9880 9890 9900
TTAAATGCTGCACCTTGCTAACATTTGATCAGTTACCGCTTCTTCAACGTGTACTTTAAGT

9910 9920 9930 9940 9950 9960
TGGTCAATCCAAGAATAACCAACACCATCAGAGAATTGTTCTTTTGACGCCATAACAATT

9970 9980 9990 10000 10010 10020
TCTTCAGGTAGTAAATCTTCAAATGCTTCTCGAATGATGTTTTCTCAATGCGGTGCGCC

10030 10040 10050 10060 10070 10080
GTGATCATTTTTAGTTCAAGGTTTAGACGCATTGACGCATCAACAAATTCTTTATCTAAG

10090 10100 10110 10120 10130 10140
AAAGGAACACGTGCTTCGATGCCCCAAGCTGCCATAGATTTGTTTGACGTAAGCAATCA

10150 10160 10170 10180 10190 10200
AACATATGTAATTTATTTACTTTACGTACCGTCTCTTCATGGAATTCTTTCGCATTGCGC

10210 10220 10230 10240 10250 10260
GCTTTGTGGAAGTACAAGTAACCACCGAACAGTTCATCAGCACCTTACCAGAAAGCACC

Fig. 5

10270 10280 10290 10300 10310 10320
ATCTTAATCCCATGGCTTTAATTTTACGTGCCATTAGGTACATAGGGGTTGATGCACGA

10330 10340 10350 10360 10370 10380
ATTGTTGTTACATCGTAGGTTTCAATGTGGTAAATCACGTGCGTAAAGCGTCGATACCT

10390 10400 10410 10420 10430 10440
TCTTGACAGTAAATTCAATTGAATGATGGATAGTACCTAAGTGATCTGCCACTTTTTTGT

10450 10460 10470 10480 10490 10500
GCAGCGGCTAAATCTGGAGAACCATTTAGGCCTACAGAGAAAGAGTGATGTTGTGGCCAC

10510 10520 10530 10540 10550 10560
CATGCTTCGGTTTTACCACCGTCTTCAATACGACGTTTTGCATACTGTTGGGTGATTGCT

10570 10580 10590 10600 10610 10620
GAAATAACAGATGAATCTAACCCGCCTGATAATAATACGCCGTAAGGTACATCACACATT

10630 10640 10650 10660 10670 10680
AATTGACGTTTAACTGCATCTTCCAAACCTTGCTTAACAACGCTTTTATCACCACCATT

10690 10700 10710 10720 10730 10740
TGTGCAACGTTATCAAATCTTTCCAATCACGTTGATAATAAGGCGTGACTACACCATCC

10750 10760 10770 10780 10790 10800
TTACTCCACAGGTAATGACCTGCTGGGAATTCTTCAATTTGAGTACAAATTGGCACTAGT

10810 10820 10830 10840 10850 10860
GCTTTTCATTTTACAGAGGCAACATAAAAGTTACCGTGTTTCATCATAGCCCGTATAAAGAGGG

10870 10880 10890 10900 10910 10920
ATGATACCGATATGGTCACGGCCAATCAGGTAAGCGTCCTCTGTTTCGTCATATAAAGCG

10930 10940 10950 10960 10970 10980
AAAGCAAAAATACCATTTAGATCATCTAAAAATTGTGTGCCTTTTTCTTTATATAGCGCA

10990 11000 11010 11020 11030 11040
AGTATCACTTCGCAATCTGATTCTGTTTGGAATTCAAAGTCTACGTTTCAGCGTTTTCTTT

11050 11060 11070 11080 11090 11100
AAATCTTTGTGGTTATAAATTTTACCATTAAACAGCAAGTACGTGTGTCTTTTCTTCATTA

11110 11120 11130 11140 11150 11160
TATAGCGGCTGTGCACCATTATTTACATCGACAATAGCAAGACGTTTCATGAACTAAAATA

11170 11180 11190 11200 11210 11220
GCATTGTCACCTTGATAGATACCTGACCAATCTGGGCCGCGGTGACGTAGTAACCTTGAT

11230 11240 11250 11260 11270 11280
AGTTCTAGTGCTTGTTCGGAAGAGGTTTAAATGTCTGATTTGATGTCTAGAATTCCGAAT

11290 11300 11310 11320 11330 11340
ATTGAGCACATAACTAATTCCTTCTGGGGCTGCGTCTGCAGCTAACTTTCTAAATAGTGT

11350 11360 11370 11380 11390 11400
GTCTAATTTGCCACATTGTAGATTTAATGCAAACATTAATGATAAAACATTTATAAAAAA

Fig. 5

11410 11420 11430 11440 11450 11460
TGTAATTC AATGTGGAATCGATAATTTAATGGCTTAAAAGTGAAGATCCATTAAATTGTGA

11470 11480 11490 11500 11510 11520
TGGCGAGGTGATAGACCAATGTAGACCTTAATGAATAAAGCAGGCACGATTGAATCCATT

11530 11540 11550 11560 11570 11580
CAACGCAAAGTGGTACTAACTATTGTTTTAAACGTTATAAATAGTGTTTTAAAGGTTATA

11590 11600 11610 11620 11630 11640
AGTAAATAATTTAAAAACAATAATAATCCACATGCATTAAATTTATCATGATAAACCGCT

11650 11660 11670 11680 11690 11700
ATATCTCAATGGCAATTTGGGATAAGTGTAATAATATATGTAAATGAATGAGTTGACTTG

11710 11720 11730 11740 11750 11760
CTTTTTTTACACTAAGTGATGAAATTAAAGCTAGATGTCGTTGTTAGCATTGATTAATAA

11770 11780 11790 11800 11810 11820
CGTACTAAAATACGACATCTAGTATAGAAATTTAAAAACAGTTGGTTTTGATAGCATAA

11830 11840 11850 11860 11870 11880
CTGCATAAACTAATCAGCTTATTGTCGTGAATATTTTTGTAATTTAAATAGGTTTAATAA

11890 11900 11910 11920 11930 11940
AATTATATGTCGTGATAAATATAAACCGTACGACCTTTCCTTTAAAAAGACGTTTTTGCTG

11950 11960 11970 11980 11990 12000
CCTAAGTTTTGGCCTGTGTGGTTCGGGGTGTTTGCAATATACTTATTAGCTTTTATGCCA

12010 12020 12030 12040 12050 12060
GTAAAGCCGCGTGATAAATTTGCTCGATTTCATAGCGAAGAAATTGTTTAGTCTAAAAATG

12070 12080 12090 12100 12110 12120
ATGGCAAAGCGTAAAAAGGTAGCAAAGATCAATTTATCTATGTGCTTCCCTGAAATGGAT

12130 12140 12150 12160 12170 12180
GATACGGAACAAGACCGTATAATCATGGTCAATCTAGTTACTTTTTGTCAAACATCTTA

12190 12200 12210 12220 12230 12240
AGTTATGCAGAGCCAAGTGC GCGTAGTCGTGCTTATAACCGTGACCGTATGATAGTGCAT

12250 12260 12270 12280 12290 12300
GGTGGCGAGAATTTATTTCCGCTACTTGAACAAGGTAAGGCTTGTATCTTATTAGTGCCG

12310 12320 12330 12340 12350 12360
CATAGCTTCGCTATTGATTTTGCAGGTTTACACATTGCTTCTTATGGCGCGCCATTTGT

12370 12380 12390 12400 12410 12420
ACTATGTTTAAACAATTCTGAGAATGAGTTGTTTCGATTGGCTGATGACACGTCAACGCGCT

12430 12440 12450 12460 12470 12480
ATGTTTGGAGGCACTGTTTATCACC GCAAGGCAGGGCTAGGGGCTCTAGTTAAATCACTT

12490 12500 12510 12520 12530 12540
AAGAGCGGTGAAAGCTGTTATTACTTACCTGATGAAGACCATGGACCTAAGCGTAGTGTA

Fig. 5

12550 12560 12570 12580 12590 12600
TTTGCGCCTTTATTTGCGACTCAAAAAGCAACTTTACCTGTAATGGGCAAGCTAGCAGAA

12610 12620 12630 12640 12650 12660
AAAACAAATGCACTCGTTGTTCCCTGTTTATGCGGCATATAATGAATCACTAGGTAAATTT

12670 12680 12690 12700 12710 12720
GAAACCTTTATTCGACCAGCAATGCAAACTTTCCATCAGAAAGCCCAGAACAAGATGCA

12730 12740 12750 12760 12770 12780
GTGATGATGAATAAAGAGATTGAAGCCTTGATTGAATGTGGTGTGATCAATATATGTGG

12790 12800 12810 12820 12830 12840
ACACTTAGATTATTGAGAACACGTCCGGACGGTAAAAAATCTACTAATAAAGTTTAATA

12850 12860 12870 12880 12890 12900
AACACCATAATCTTCGTTGAATATGGTGTTCACCCCTGAATACCTCTAAATTAATAA

12910 12920 12930 12940 12950 12960
CAAAAAAGCCATTTACGTAACATCTAATGATGATTTAGCCTGCACTTGCTTTGTTTTTA

12970 12980 12990 13000 13010 13020
GTCTTAAGAGCCTAATAAACTTGATCTAGGTATAGATTCTGTCTTTCTTTACGTAACGCG

13030 13040 13050 13060 13070 13080
ATCTATTTTTTTTAACCGATAGTTGTTATAATTAGTTTCATATGAAAGAGATATCGTTTC

13090 13100 13110 13120 13130 13140
AGTAAAAGCTATTTTCGTTTCAATAGATAATTTATTTATAGTCATATTTCTGTAATGACA

13150 13160 13170 13180 13190 13200
ATCATTTTCTCATCTAGACTATAGATAAGAATACGAATTAAGTAAGAACATTAATTTTAC

13210 13220 13230 13240 13250 13260
AAGAATATAAAATATCCCATCGGAGCTATAAGAATGAAAAAGACTAAAATTGTTTGTACA

13270 13280 13290 13300 13310 13320
ATTGGTCCAAAACTGAATCAGTAGAGAACTAACAGAGCTTGTTAATGCAGGCATGAAC

13330 13340 13350 13360 13370 13380
GTTATGCGTTTAAATTTCTCTCATGGTAACTTTGCTGAACATTCAGTGCGTATTCAAAT

13390 13400 13410 13420 13430 13440
ATCCGTCAAGTAAGTGAACCTGAATAAGAAAATTGCTGTTTTACTGGATACTAAAGGT

13450 13460 13470 13480 13490 13500
CCAGAAATCCGTACGATTAACTAGAAAACGGTGACGATGTAATGTTGACCGCTGGTCAG

13510 13520 13530 13540 13550 13560
TCATTACGTTTACAACAGACATTAACGTGGTAGGTAATAAAGACTGTGTTGCTGTAACA

13570 13580 13590 13600 13610 13620
TATGCTGGTTTTGCTAAAGACCTTAATCCTGGTGCAATCATCCTTGTTGATGATGGTTTA

13630 13640 13650 13660 13670 13680
ATTGAAATGGAAGTTGTTGCAACAACTGACACTGAAGTTAAATGTACAGTATTAAATACT

Fig. 5

13690 13700 13710 13720 13730 13740
GGTGCACCTTGGTGAAAATAAAGGCGTTAACTTACCTAACATCAGTGTAGGTCTACCTGCA

13750 13760 13770 13780 13790 13800
TTGTCAGAAAAAGATAAAGCTGATTTAGCGTTTGGTTGTGAGCAAGAAGTTGATTTTGGTT

13810 13820 13830 13840 13850 13860
GCTGCATCATTTATTTCGTAAGGCTGATGATGTAAGAGAAATTCGTGAAATCCTATTTAAT

13870 13880 13890 13900 13910 13920
AATGGTGGCGAAAAACATTTCAGATTATCTCGAAAATTGAAAACCAAGAAGGTGTAGACAAT

13930 13940 13950 13960 13970 13980
TTCGATGAAATCTTAGCTGAATCAGACGGTATCATGGTTGCTCGTGGCGATCTCGGTGTT

13990 14000 14010 14020 14030 14040
GAGATCCCAGTTGAAGAAGTGATCATGGCACAGAAGATGATGATCAAAAAATGTAATAAA

14050 14060 14070 14080 14090 14100
GCAGGTAAAGTTGTAATTACTGCAACACAAATGCTTGATTCAATGATCAGTAACCCACGT

14110 14120 14130 14140 14150 14160
CCAACACGTGCAGAAGCGGGCGATGTTGCCAATGCTGTGCTTGACGGTACCGACGCGGTA

14170 14180 14190 14200 14210 14220
ATGCTTTCTGGTGAACTGCGAAAGGTAAATACCCAGTTGAAGCTGTGTCTATCATGGCA

14230 14240 14250 14260 14270 14280
AACATCTGTGAACGTACTGATACTCAATGTCTTCGGATTTAGGTGCGAACATTGTTGCT

14290 14300 14310 14320 14330 14340
AAAAGCATGCGCATTACAGAAGCTGTGTGTAAAGGTGCGGTAGAAACAACAGAAAAATTG

14350 14360 14370 14380 14390 14400
TGTGCTCCACTTATTGTTGTTGCAACTCGTGGCGGTAAATCAGCAAAATCTGTTTCGTAAA

14410 14420 14430 14440 14450 14460
TACTTCCCGAAAGCAAATATTCTTGCTATCACAACAAATGAAAAAGCAGCGCAACAGTTA

14470 14480 14490 14500 14510 14520
TGCCTAACTAAAGGCGTAAGCAGCTGCATCGTTGAGCAGATTGATAGCACTGATGAGTTC

14530 14540 14550 14560 14570 14580
TACCGTAAAGGTAAAGAGCTTGATTAGCAACTGGTTTAGCTAAAGAAGGCGATATCGTT

14590 14600 14610 14620 14630 14640
GTTATGGTATCAGGTGCGTTAGTACCATCAGGTACAACGAATACGGCATCTGTTACCAA

14650 14660 14670 14680 14690 14700
CTTTAAGTTGCCATATTGATATTATAAAAAAGAGAGCGTATGCTCTCTTTTTTTATATCT

14710 14720 14730 14740 14750 14760
GTAGTTTATATGTCTGTACAAAAAATGATAAAGAGTACATAAACTATTAATATAGCGTA

14770 14780 14790 14800 14810 14820
ATATATAATGATTAACGGTGATGAAAGGGTTAAATAAATGGATAGTGCTAAACATAAAAT

Fig. 5

14830 14840 14850 14860 14870 14880
TGGCTTAGTCCTTTCTGGCGGTGGTGCAGAAAGGTATTGCTCATCTTGGTGTATTAAAAATA

14890 14900 14910 14920 14930 14940
CCTGTTAGAGCAAGATATAAGACCGAATGTAATTGCGGGTACAAGTGCTGGCTCTATGGT

14950 14960 14970 14980 14990 15000
TGGTGCACTTTATTGCTCAGGACTTGAGATTGATGACATTTTACAATTCTTCATCGATGT

15010 15020 15030 15040 15050 15060
AAAACCTTTTTCTTGGAAGTTTACCCGTGCCCGTGCTGGCTTTATAGACCCGGCAAAT

15070 15080 15090 15100 15110 15120
ATATCCTGAAGTGCTAAAATATATCCCCGAGGATAGCTTTGAGTACCTTCAACCTGAATT

15130 15140 15150 15160 15170 15180
GCGCATTGTTGCCACCAACATGTTACTCGGTAAAGAGCATATATTTAAAGATGGCTCCGT

15190 15200 15210 15220 15230 15240
GATTAATGCCTTATTAGCATCAGCCAGCTACCCTTTAGTTTTTCTCCGATGATCATTGA

15250 15260 15270 15280 15290 15300
CGATCAAGTGTATTGATGGCGGTATTGTTAATCATTTCCCGTGAGTGTATTGAAGA

15310 15320 15330 15340 15350 15360
TGATTGCGATAAAATAATCGGCGTATACGTGTCGCCCATTTCGTCAGGTGCAAGCTGACGA

15370 15380 15390 15400 15410 15420
ACTCTCGAGTATAAAAGACGTGGTATTACGTGCGTTTACGCTGCAGGGTAGTGGTGCTGA

15430 15440 15450 15460 15470 15480
ATTAGATAAACTATCGCAATGTGATGTGCAAATTTATCCAGAAGCGCTATTGAATTACAA

15490 15500 15510 15520 15530 15540
TACGTTTGCAACCGATGAAAAATCATTACGGGAGATCTACCAGATTGGTTATGATGCTGC

15550 15560 15570 15580 15590 15600
AAAAGATCAACATGACAACCTTATGGCATTGAAAGAAAGTATCACCACCAGCGAGGTTAA

15610 15620 15630 15640 15650 15660
AAAGAACGTCTTTAGCAAATGGTTTGGTGATAAACTTGCTAGCAACAGCGGCAAATAGCG

15670 15680 15690 15700 15710 15720
GCCCCACACGATTTATACACTAGGATAATGGGCGTTAATAGCCTCACTGTCGTTGTGTGG

15730 15740 15750 15760 15770 15780
TCTCTAATTTTAGCTAAATCTTGTGTTATACTGACTTCCTATTAATCATAAACGATTTAT

15790 15800 15810 15820 15830 15840
CACGGTAAACATGACTCAAATAAATAACCCGCTTCACGGCATGACACTCGAAAAAGTAAT

15850 15860 15870 15880 15890 15900
TAACAGTCTCGTTGAACAATATGGCTGGGATGGTCTTGGATACTACATCAACATTCTGTTG

15910 15920 15930 15940 15950 15960
CTTTACTGAAAATCCAAGTGTTAAGTCTAGTCTTAAATTTTTACGTAAAACCCCTTGGGC

Fig. 5

AM

15970 15980 15990 16000 16010 16020
ACGTGATAAAGTAGAAGCGCTATATATCAAATGGTGACTGAAGGCTAACTGTCTCCACG

16030 16040 16050 16060 16070 16080
CTAGCGAACCGCTGTTTATAGTTAATATAAGTACTATAAGCAGGGCTCGTTAATTCAGTA

16090 16100 16110 16120 16130 16140
TGTAATTAATCCTGAATACCTCCGCTTATTTCAACATTGTACTCTCTAGATAACACTCTC

16150 16160 16170 16180 16190 16200
AACATTACACCTTCAACATCACAGCCTCCACATAACATCCGATGACATAGCCCTGTTATT

16210 16220 16230 16240 16250 16260
TTTCACATTTATCTATATGCTATATATTTTAGCCATTTGATCAATTGAGTTAATTTCTGC

16270 16280 16290 16300 16310 16320
AATGACAAAGATATACCATCATCCAGTACAAATTTATTATGAAGATACCGACCATTCTGG

16330 16340 16350 16360 16370 16380
TGTTGTTTACCACCCTAACTTTTTAAATACTTTGAACGTGCACGTGAGCATGTGATAAA

16390 16400 16410 16420 16430 16440
TAGTGACTTACTAGCAACATTGTGGAATGAACGCGGTTTAGGTTTTGCGGTGTATAAAGC

16450 16460 16470 16480 16490 16500
CAATATGACTTTTCAGGATGGGGTCGAATTTGCTGAAGTGTGTGATATTCGCACTTCTTT

16510 16520 16530 16540 16550 16560
TGTCTAGACGGTAAGTACAAAACGATCTGGCGCCAAGAAGTATGGCGTCCGAATGCGAC

16570 16580 16590 16600 16610 16620
TAGGGCTGCCGTTATCGGTGATATTGAAATGGTGTGCTTAGACAAACAAAACGTTTACA

16630 16640 16650 16660 16670 16680
GCCCATCCCTGATGATGTGTTAGCTGCAATGGTTAGTGAATAAATGGTTCATGCATAAAT

16690 16700 16710 16720 16730 16740
AGTTAATACATGATTCTGGCCCGTCACGTTTACAGATAAGAGGCATCCGATGCCTCCTTC

16750 16760 16770 16780 16790 16800
CTATTACCAATACTACTGCTTATCCCTTTCTAACTATCTTTAGCGTCCATAACACACTGA

16810 16820 16830 16840 16850 16860
GCATTTATTCTATTAATCAGTGATTGTGATTTAATTATCTTCTATATATGTAATTTAATG

16870 16880 16890 16900 16910 16920
TAATTTTCAATTTATTTTATAGCTACATTAAGGCTTACGAATGTACGCTAAAATGAGATGT

16930 16940 16950 16960 16970 16980
CAGACTAATTTTAGCTTATTAATCTGTTAGCCGTTTATATTTTATAAAGATGGGATTTAA

16990 17000 17010 17020 17030 17040
CTTAAATGCAATTAATTATGGCGTAAATAGAGTGAAAACATGGCTAATATTCACCTAAGTC

17050 17060 17070 17080 17090 17100
CTGAATTTTATATAAAGTTTAATCTGTTATTTTAGCGTTTACCTGGTCTTATCAGTGAGG

Fig. 5

17110 17120 17130 17140 17150 17160
TTTATAGCCATTATTAGTGGGATTGAAGTGATTTTTAAAGCTATGTATATTATTGCAAAT

17170 17180 17190 17200 17210 17220
ATAAATTGTAACAATTAAGACTTTGGACACTTGAGTTCAATTCGAATTGATTGGCATAA

17230 17240 17250 17260 17270 17280
AATTTAAACAGCTAAATCTACCTCAATCATTTTAGCAAATGTATGCAGGTAGATTTTTT

17290 17300 17310 17320 17330 17340
TCGCCATTTAAGAGTACACTTGTACGCTAGGTTTTTGTGTTAGTGTGCAAATGAACGTTTT

17350 17360 17370 17380 17390 17400
GATGAGCATTGTTTTTAGAGCACAAAATAGATCCTTACAGGAGCAATAACGCAATGGCTA

17410 17420 17430 17440 17450 17460
AAAAGAACCACATCGATTAAGCACGCCAAGGATGTGTTAAGTAGTGATGATCAACAGT

17470 17480 17490 17500 17510 17520
TAAATTCTCGCTTGCAAGAATGTCCGATTGCCATCATTGGTATGGCATCGGTTTTTGCAG

17530 17540 17550 17560 17570 17580
ATGCTAAAACTTGGATCAATTCTGGGATAACATCGTTGACTCTGTGGACGCTATTATTG

17590 17600 17610 17620 17630 17640
ATGTGCCTAGCGATCGCTGGAACATTGACGACCATTACTCGGCTGATAAAAAAGCAGCTG

17650 17660 17670 17680 17690 17700
ACAAGACATACTGCAAACGCGGTGGTTTCATTCCAGAGCTTGATTTTGATCCGATGGAGT

17710 17720 17730 17740 17750 17760
TTGGTTTACCGCCAAATATCCTCGAGTTAACTGACATCGCTCAATTGTTGTCAATTAATTG

17770 17780 17790 17800 17810 17820
TTGCTCGTGATGTATTAAGTGATGCTGGCATTGGTAGTGATTATGACCATGATAAAATTG

17830 17840 17850 17860 17870 17880
GTATCAGCTGGGTGTCGGTGGTGGTCAGAAACAAATTCGCCATTAACGTGCGCGCTAC

17890 17900 17910 17920 17930 17940
AAGGCCCGGTATTAGAAAAAGTATTAAAAGCCTCAGGCATTGATGAAGATGATCGCGCTA

17950 17960 17970 17980 17990 18000
TGATCATCGACAAATTTAAAAAGCCTACATCGGCTGGGAAGAGAACTCATTCCCAGGCA

18010 18020 18030 18040 18050 18060
TGCTAGGTAACGTTATTGCTGGTTCGTATCGCCAATCGTTTTGATTTTGGTGGTACTAACT

18070 18080 18090 18100 18110 18120
GTGTGGTTGATGCGGCATGCGCTGGCTCCCTTGACAGCTGTTAAATGGCGATCTCAGACT

18130 18140 18150 18160 18170 18180
TACTTGAATATCGTTTCAAGTCATGATATCGGGTGGTGTATGTTGTGATAACTCGCCAT

18190 18200 18210 18220 18230 18240
TCATGTATATGTCATTCTCGAAAACACCAGCAATTACCACCAATGATGATATCCGTCCGT

Fig. 5

18250 18260 18270 18280 18290 18300
TTGATGACGATTCAAAGGCATGCTGGTTGGTGAAGGTATTGGCATGATGGCGTTTAAAC

18310 18320 18330 18340 18350 18360
GTCTTGAAGATGCTGAACGTGACGGCGACAAAATTTATTCTGTACTGAAAGGTATCGGTA

18370 18380 18390 18400 18410 18420
CATCTTCAGATGGTCGTTTCAAATCTATTTACGCTCCACGCCAGATGGCCAAGCAAAAG

18430 18440 18450 18460 18470 18480
CGCTAAAACGTGCTTATGAAGATGCCGGTTTTGCCCTGAAACATGTGGTCTAATTGAAG

18490 18500 18510 18520 18530 18540
GCCATGGTACGGGTACCAAAGCGGGTGATGCCGCAGAATTTGCTGGCTTGACCAAACT

18550 18560 18570 18580 18590 18600
TTGGCGCCGCCAGTGATGAAAAGCAATATATCGCCTTAGGCTCAGTTAAATCGCAAATTG

18610 18620 18630 18640 18650 18660
GTCATACTAAATCTGCGGCTGGCTCTGCGGGTATGATTAAGGCGGCATTAGCGCTGCATC

18670 18680 18690 18700 18710 18720
ATAAAATCTTACCTGCAACGATCCATATCGATAAACCAAGTGAAGCCTTGGATATCAAAA

18730 18740 18750 18760 18770 18780
ACAGCCCGTTATACCTAAACAGCGAAACGCGTCCTTGGATGCCACGTGAAGATGGTATTC

18790 18800 18810 18820 18830 18840
CACGTCGTGCAGGTATCAGCTCATTTGGTTTTGGCGGCACCAACTTCCATATTATTTTAG

18850 18860 18870 18880 18890 18900
AAGAGTATCGCCCAGGTACGATAGCGCATATCGCTTAAACTCAGTGAGCCAACTGTGT

18910 18920 18930 18940 18950 18960
TGATCTCGGCAAACGACCAACAAGGTATTGTTGCTGAGTTAAATAACTGGCGTACTAAAC

18970 18980 18990 19000 19010 19020
TGGCTGTGATGCTGATCATCAAGGGTTTGTATTTAATGAGTTAGTGACAACGTGGCCAT

19030 19040 19050 19060 19070 19080
TAAAAACCCCATCCGTTAACCAAGCTCGTTTTAGGTTTTGTTGCGCGTAATGCAAATGAAG

19090 19100 19110 19120 19130 19140
CGATCGCGATGATTGATACGGCATTGAAACAATTCAATGCGAACGCAGATAAAATGACAT

19150 19160 19170 19180 19190 19200
GGTCAGTACCTACCGGGGTTTACTATCGTCAAGCCGGTATTGATGCAACAGGTAAAGTGG

19210 19220 19230 19240 19250 19260
TTGCGCTATTCTCAGGGCAAGGTTTCGCAATACGTGAACATGGGTCGTGAATTAACCTGTA

19270 19280 19290 19300 19310 19320
ACTTCCCAAGCATGATGCACAGTGCTGCGGCGATGGATAAAGAGTTCAGTGCCGCTGGTT

19330 19340 19350 19360 19370 19380
TAGGCCAGTTATCTGCAGTTACTTTCCCTATCCCTGTTTATACGGATGCCGAGCGTAAGC

Fig. 5

19390 19400 19410 19420 19430 19440
TACAAGAAGAGCAATTACGTTTAAACGCAACATGCGCAACCAGCGATTGGTAGTTTGAGTG

19450 19460 19470 19480 19490 19500
TTGGTCTGTTCAAAACGTTTAAAGCAAGCAGGTTTTAAAGCTGATTTTGCTGCCGGTCATA

19510 19520 19530 19540 19550 19560
GTTTCGGTGAGTTAACCGCATTATGGGCTGCCGATGTATTGAGCGAAAGCGATTACATGA

19570 19580 19590 19600 19610 19620
TGTTAGCGCGTAGTCGTGGTCAAGCAATGGCTGCGCCAGAGCAACAAGATTTTGATGCAG

19630 19640 19650 19660 19670 19680
GTAAGATGCCCGCTGTTGTTGGTGATCCAAAGCAAGTCGCTGTGATCATTGATACCCTTG

19690 19700 19710 19720 19730 19740
ATGATGTCTCTATTGCTAACTTCAACTCGAATAACCAAGTTGTTATTGCTGGTACTACGG

19750 19760 19770 19780 19790 19800
AGCAGGTTGCTGTAGCGGTTACAACCTTAGGTAATGCTGGTTTCAAAGTTGTGCCACTGC

19810 19820 19830 19840 19850 19860
CGGTATCTGCTGCGTTCATACACCTTTAGTTCGTACGCGCAAAAACCATTTGCTAAAG

19870 19880 19890 19900 19910 19920
CGGTTGATAGCGCTAAATTTAAAGCGCCAAGCATTCCAGTGTTTGCTAATGGCACAGGCT

19930 19940 19950 19960 19970 19980
TGGTGCAATCAAGCAAACCGAATGACATTAAGAAAAACCTGAAAAACCACATGCTGGAAT

19990 20000 20010 20020 20030 20040
CTGTTCAATTTCAATCAAGAAATTGACAACATCTATGCTGATGGTGCCCGCTATTTATCG

20050 20060 20070 20080 20090 20100
AATTTGGTCCAAAGAATGTATTAATACTAAATTGGTTGAAAACATTCTCACTGAAAAATCTG

20110 20120 20130 20140 20150 20160
ATGTGACTGCTATCGCGGTTAATGCTAATCCTAAACAACCTGCGGACGTACAAATGCGCC

20170 20180 20190 20200 20210 20220
AAGCTGCGCTGCAAATGGCAGTGCTTGGTGTGCGATTAGACAATATTGACCCGTACGACG

20230 20240 20250 20260 20270 20280
CCGTTAAGCGTCCACTTGTTGCGCCGAAAGCATCACCAATGTTGATGAAGTTATCTGCAG

20290 20300 20310 20320 20330 20340
CGTCTTATGTTAGTCCGAAAACGAAGAAAGCGTTTGCTGATGCATTGACTGATGGCTGGA

20350 20360 20370 20380 20390 20400
CTGTTAAGCAAGCGAAAGCTGTACCTGCTGTTGTGTCACAACCACAAGTGATTGAAAAGA

20410 20420 20430 20440 20450 20460
TCGTTGAAGTTGAAAAGATAGTTGAACGCATTGTGCAAGTAGAGCGTATTGTGCAAGTAG

20470 20480 20490 20500 20510 20520
AAAAAATCGTCTACGTTAATGCTGACGGTTCGCTTATATCGCAAAATAATCAAGACGTTA

Fig. 5

20530 20540 20550 20560 20570 20580
ACAGCGCTGTTGTTAGCAACGTGACTAATAGCTCAGTGACTCATAGCAGTGATGCTGACC

20590 20600 20610 20620 20630 20640
TTGTTGCCTCTATTGAACGCAGTGTGGTCAATTTGTTGCACACCAACAGCAATTATTAA

20650 20660 20670 20680 20690 20700
ATGTACATGAACAGTTTATGCAAGGTCCACAAGACTACGCGAAAACAGTGCAGAACGTAC

20710 20720 20730 20740 20750 20760
TTGCTGCGCAGACGAGCAATGAATTACCGGAAAGTTTAGACCGTACATTGTCTATGTATA

20770 20780 20790 20800 20810 20820
ACGAGTTCCAATCAGAAACGCTACGTGTACATGAAACGTACCTGAACAATCAGACGAGCA

20830 20840 20850 20860 20870 20880
ACATGAACACCATGCTTACTGGTGCTGAAGCTGATGTGCTAGCAACCCCAATAACTCAGG

20890 20900 20910 20920 20930 20940
TAGTGAATACAGCCGTTGCCACTAGTCACAAGGTAGTTGCTCCAGTTATTGCTAATACAG

20950 20960 20970 20980 20990 21000
TGACGAATGTTGTATCTAGTGTGAGTAATAACGCGGCGGTTGCAGTGCAAACGTGTGGCAT

21010 21020 21030 21040 21050 21060
TAGCGCTACGCAAGAAATCGCTCCAACAGTCGCTACTACGCCAGCACCCGCATTGGTTG

21070 21080 21090 21100 21110 21120
CTATCGTGGCTGAACCTGTGATTGTTGCGCATGTTGCTACAGAAGTTGCACCAATTACAC

21130 21140 21150 21160 21170 21180
CATCAGTTACACCAGTTGTGCAACTCAAGCGGCTATCGATGTAGCAACTATTAACAAAG

21190 21200 21210 21220 21230 21240
TAATGTTAGAAGTTGTTGCTGATAAAACCGGTTATCCAACGGATATGCTGGAACGTAGCA

21250 21260 21270 21280 21290 21300
TGGACATGGAAGCTGACTTAGGTATCGACTCAATCAAACGTGTTGAGATATTAGGCGCAG

21310 21320 21330 21340 21350 21360
TACAGGAATTGATCCCTGACTTACCTGAACTTAATCCTGAAGATCTTGCTGAGCTACGCA

21370 21380 21390 21400 21410 21420
CGCTTGGTGAGATTGTCGATTACATGAATTCAAAAGCCCAGGCTGTAGCTCCTACAACAG

21430 21440 21450 21460 21470 21480
TACCTGTAACAAGTGCACCTGTTTCGCCTGCATCTGCTGGTATTGATTTAGCCACATCC

21490 21500 21510 21520 21530 21540
AAAACGTAATGTTAGAAGTGGTTGCAGACAAAACCGGTTACCCAACAGACATGCTAGAAC

21550 21560 21570 21580 21590 21600
TGAGCATGGATATGGAAGCTGACTTAGGTATTGATTCAATCAAGCGTGTGGAAATCTTAG

21610 21620 21630 21640 21650 21660
GTGCAGTACAGGAGATCATAACTGATTTACCTGAGCTAAACCCTGAAGATCTTGCTGAAT

Fig. 5

21670 21680 21690 21700 21710 21720
TACGCACCCCTAGGTGAAATCGTTAGTTACATGCAAAGCAAAGCGCCAGTCGCTGAAAGTG

21730 21740 21750 21760 21770 21780
CGCCAGTGGCGACGGCTCCTGTAGCAACAAGCTCAGCACCGTCTATCGATTTGAACCACA

21790 21800 21810 21820 21830 21840
TTCAAACAGTGATGATGGATGTAGTTGCAGATAAGACTGGTTATCCAACGTGACATGCTAG

21850 21860 21870 21880 21890 21900
AACTTGGCATGGACATGGAAGCTGATTTAGGTATCGATTCAATCAAACGTGTGGAAATAT

21910 21920 21930 21940 21950 21960
TAGGCGCAGTGCAGGAGATCATCACTGATTTACCTGAGCTAAACCCAGAAGACCTCGCTG

21970 21980 21990 22000 22010 22020
AATTACGCACGCTAGGTGAAATCGTTAGTTACATGCAAAGCAAAGCGCCAGTCGCTGAGA

22030 22040 22050 22060 22070 22080
GTGCGCCAGTAGCGACGGCTTCTGTAGCAACAAGCTCTGCACCGTCTATCGATTTAAACC

22090 22100 22110 22120 22130 22140
ATATCCAAACAGTGATGATGGAAGTGGTTGCAGACAAAACCGGTTATCCAGTAGACATGT

22150 22160 22170 22180 22190 22200
TAGAACTTGCTATGGACATGGAAGCTGACCTAGGTATCGATTCAATCAAGCGTGTAGAAA

22210 22220 22230 22240 22250 22260
TTTTAGGTGCGGTACAGGAAATCATTACTGACTTACCTGAGCTTAACCCTGAAGATCTTG

22270 22280 22290 22300 22310 22320
CTGAACTACGTACATTAGGTGAAATCGTTAGTTACATGCAAAGCAAAGCGCCCGTAGCTG

22330 22340 22350 22360 22370 22380
AAGCGCCTGCAGTACCTGTTGCAGTAGAAAGTGCACCTACTAGTGTAACAAGCTCAGCAC

22390 22400 22410 22420 22430 22440
CGTCTATCGATTTAGACCACATCCAAAATGTAATGATGGATGTTGTTGCTGATAAGACTG

22450 22460 22470 22480 22490 22500
GTTATCCTGCCAATATGCTTGAATTAGCAATGGACATGGAAGCCGACCTTGGTATTGATT

22510 22520 22530 22540 22550 22560
CAATCAAGCGTGTGAAATTCTAGGCGCGGTACAGGAGATCATTACTGATTTACCTGAAC

22570 22580 22590 22600 22610 22620
TAAACCCAGAAGACTTAGCTGAACTACGTACGTTAGAAGAAATTGTAACCTACATGCAA

22630 22640 22650 22660 22670 22680
GCAAGGCGAGTGGTGTTACTGTAAATGTAGTGGCTAGCCCTGAAAATAATGCTGTATCAG

22690 22700 22710 22720 22730 22740
ATGCATTTATGCAAAGCAATGTGGCGACTATCACAGCGGCCGAGAACATAAGGCGGAAT

22750 22760 22770 22780 22790 22800
TTAAACCGGCGCCGAGCGCAACCGTTGCTATCTCTCGTCTAAGCTCTATCAGTAAAATAA

Fig. 5

22810 22820 22830 22840 22850 22860
GCCAAGATTGTAAAGGTGCTAACGCCTTAATCGTAGCTGATGGCACTGATAATGCTGTGT

22870 22880 22890 22900 22910 22920
TACTTGCAGACCACCTATTGCAAACCTGGCTGGAATGTAACCTGCATTGCAACCAACTTGGG

22930 22940 22950 22960 22970 22980
TAGCTGTAACAACGACGAAAGCATTTAATAAGTCAGTGAACCTGGTGACTTTAAATGGCG

22990 23000 23010 23020 23030 23040
TTGATGAAACTGAAATCAACAACATTATTACTGCTAACGCACAATTGGATGCAGTTATCT

23050 23060 23070 23080 23090 23100
ATCTGCACGCAAGTAGCGAAATTAATGCTATCGAATACCCACAAGCATCTAAGCAAGGCC

23110 23120 23130 23140 23150 23160
TGATGTTAGCCTTCTTATTAGCGAAATTGAGTAAAGTAACTCAAGCCGCTAAAGTGCCTG

23170 23180 23190 23200 23210 23220
GCGCCTTTATGATTGTTACTCAGCAGGGTGGTTCATTAGGTTTTGATGATATCGATTCTG

23230 23240 23250 23260 23270 23280
CTACAAGTCATGATGTGAAACAGACCTAGTACAAAGCGGCTTAAACGGTTTAGTTAAGA

23290 23300 23310 23320 23330 23340
CACTGTCTCACGAGTGGGATAACGTATTCTGTCGTGCGGTTGATATTGCTTCGTCATTAA

23350 23360 23370 23380 23390 23400
CGGCTGAACAAGTTGCAAGCCTTGTTAGTGATGAACTACTTGATGCTAACACTGTATTAA

23410 23420 23430 23440 23450 23460
CAGAAGTGGGTTATCAACAAGCTGGTAAAGGCCCTGAACGTATCACGTTAACTGGTGTGG

23470 23480 23490 23500 23510 23520
CTACTGACAGCTATGCATTAACAGCTGGCAATAACATCGATGCTAACTCGGTATTTTTAG

23530 23540 23550 23560 23570 23580
TGAGTGGTGGCGCAAAGGTGTAACCTGCACATTGTGTTGCTCGTATAGCTAAAGAATATC

23590 23600 23610 23620 23630 23640
AGTCTAAGTTCATCTTATTGGGACGTTCAACGTTCTCAAGTGACGAACCGAGCTGGGCAA

23650 23660 23670 23680 23690 23700
GTGGTATTACTGATGAAGCGGCGTTAAAGAAAGCAGCGATGCAGTCTTTGATTACAGCAG

23710 23720 23730 23740 23750 23760
GTGATAAACCAACACCCGTTAAGATCGTACAGCTAATCAAACCAATCCAAGCTAATCGTG

23770 23780 23790 23800 23810 23820
AAATTGCGCAAACCTTGCTCTGCAATTACCGCTGCTGGTGGCCAAGCTGAATATGTTTCTG

23830 23840 23850 23860 23870 23880
CAGATGTAACCTAATGCAGCAAGCGTACAAATGGCAGTCGCTCCAGCTATCGCTAAGTTCTG

23890 23900 23910 23920 23930 23940
GTGCAATCACTGGCATCATTCATGGCGCGGGTGTGTTAGCTGACCAATTCATTGAGCAAA

Fig. 5

23950 23960 23970 23980 23990 24000
AAACACTGAGTGATTTTGAGTCTGTTTACAGCACTAAAATTGACGGTTTGTTATCGCTAC

24010 24020 24030 24040 24050 24060
TATCAGTCACTGAAGCAAGCAACATCAAGCAATTGGTATTGTTCTCGTCAGCGGCTGGTT

24070 24080 24090 24100 24110 24120
TCTACGGTAACCCCGCCAGTCTGATTACTCGATTGCCAATGAGATCTTAAATAAAACCG

24130 24140 24150 24160 24170 24180
CATACCGCTTTAAATCATTGCACCCACAAGCTCAAGTATTGAGCTTTAACTGGGGTCCTT

24190 24200 24210 24220 24230 24240
GGGACGGTGGCATGGTAACGCCTGAGCTTAAACGTATGTTTGACCAACGTGGTGTTTACA

24250 24260 24270 24280 24290 24300
TTATTCCACTTGATGCAGGTGCACAGTTATTGCTGAATGAACTAGCCGCTAATGATAACC

24310 24320 24330 24340 24350 24360
GTTGTCCACAAATCCTCGTGGGTAATGACTTATCTAAAGATGCTAGCTCTGATCAAAAGT

24370 24380 24390 24400 24410 24420
CTGATGAAAAGAGTACTGCTGTAAAAAGCCACAAGTTAGTCGTTTATCAGATGCTTTAG

24430 24440 24450 24460 24470 24480
TAACTAAAAGTATCAAAGCGACTAACAGTAGCTCTTTATCAAACAAGACTAGTGCTTTAT

24490 24500 24510 24520 24530 24540
CAGACAGTAGTGCTTTTCAGGTTAACGAAAACCACTTTTGTAGCTGACCACATGATCAAAG

24550 24560 24570 24580 24590 24600
GCAATCAGGTATTACCAACGGTATGCGCGATTGCTTGGATGAGTGATGCAGCAAAAGCGA

24610 24620 24630 24640 24650 24660
CTTATAGTAACCGAGACTGTGCATTGAAGTATGTCGGTTTCGAAGACTATAAATTGTTTA

24670 24680 24690 24700 24710 24720
AAGGTGTGGTTTTTGTATGGCAATGAGGCGGCGGATTACCAAATCCAATTGTGCGCCTGTGA

24730 24740 24750 24760 24770 24780
CAAGGGCGTCAGAACAGGATTCTGAAGTCCGTATTGCCGCAAAGATCTTTAGCCTGAAAA

24790 24800 24810 24820 24830 24840
GTGACGGTAAACCTGTGTTTCATTATGCAGCGACAATATTGTTAGCAACTCAGCCACTTA

24850 24860 24870 24880 24890 24900
ATGCTGTGAAGGTAGAACTTCCGACATTGACAGAAAGTGTTGATAGCAACAATAAAGTAA

24910 24920 24930 24940 24950 24960
CTGATGAAGCACAAAGCGTTATACAGCAATGGCACCTTGTTCCACGGTGAAAGTCTGCAGG

24970 24980 24990 25000 25010 25020
GCATTAAGCAGATATTAAGTTGTGACGACAAGGGCCTGCTATTGGCTTGTCAGATAACCG

25030 25040 25050 25060 25070 25080
ATGTTGCAACAGCTAAGCAGGGATCCTTCCCGTTAGCTGACAACAATATCTTGCCAATG

Fig. 5

26230 26240 26250 26260 26270 26280
TCGCATTGTCTGAAAGCCAAGGTTTAAATGTCTGCTTATCATCATACGCAAACCTTTGCATA

26290 26300 26310 26320 26330 26340
CTGCATTAAAGCAGTGCCCGTAGTGTGACTGGTGAAGGCGGGTGTTCACAGGTCGCAG

26350 26360 26370 26380 26390 26400
GTTTATTGAAATGTGTAATTGGTTTACATCAACGTTATATTCCGGCGATTAAAGATTGGC

26410 26420 26430 26440 26450 26460
AACCAACCGAGTGACAATCAAATGTCACGGTGGCGGAATTCACCATTCTATATGCCTGTAG

26470 26480 26490 26500 26510 26520
ATGCTCGACCTTGGTTCCCATGCTGATGGCTCTGCACACATTGCCGCTTATAGTTGTG

26530 26540 26550 26560 26570 26580
TGACTGCTGACAGCTATTGTCATATTCTTTTACAAGAAAACGTCTTACAAGAACTTGTTT

26590 26600 26610 26620 26630 26640
TGAAAGAAACAGTCTTGCAAGATAATGACTTAACTGAAAGCAAGCTTCAGACTCTTGAAC

26650 26660 26670 26680 26690 26700
AAAACAATCCAGTAGCTGATCTGCGCACTAATGGTTACTTTGCATCGAGCGAGTTAGCAT

26710 26720 26730 26740 26750 26760
TAATCATAGTACAAGGTAATGACGAAGCACAAATTACGCTGTGAATTAGAACTATTACAG

26770 26780 26790 26800 26810 26820
GGCAGTTAAGTACTACTGGCATAAGTACTATCAGTATTAAACAGATCGCAGCAGACTGTT

26830 26840 26850 26860 26870 26880
ATGCCCGTAATGATACTAACAAAGCCTATAGCGCAGTGCTTATTGCCGAGACTGCTGAAG

26890 26900 26910 26920 26930 26940
AGTTAAGCAAAGAAATAACCTTGGCGTTTGCTGGTATCGCTAGCGTGTTTAAATGAAGATG

26950 26960 26970 26980 26990 27000
CTAAAGAATGGAAAACCCCGAAGGGCAGTTATTTTACCGCGCAGCCTGCAAATAAACAGG

27010 27020 27030 27040 27050 27060
CTGCTAACAGCACACAGAATGGTGTACCTTCATGTACCCAGGTATTGGTGCTACATATG

27070 27080 27090 27100 27110 27120
TTGGTTTAGGGCGTGATCTATTTTCATCTATTCCCACAGATTTATCAGCCTGTAGCGGCTT

27130 27140 27150 27160 27170 27180
TAGCCGATGACATTGGCGAAAGTCTAAAAGATACTTTACTTAATCCACGCAGTATTAGTC

27190 27200 27210 27220 27230 27240
GTCATAGCTTTAAAGAACTCAAGCAGTTGGATCTGGACCTGCGCGGTAACCTTAGCCAATA

27250 27260 27270 27280 27290 27300
TCGCTGAAGCCGGTGTGGGTTTGCTTGTTGTTTACCAAGGTATTTGAAGAAGTCTTTG

27310 27320 27330 27340 27350 27360
CCGTTAAAGCTGACTTTGCTACAGGTTATAGCATGGGTGAAGTAAGCATGTATGCAGCAC

Fig. 5

27370 27380 27390 27400 27410 27420
TAGGCTGCTGGCAGCAACCGGATTGATGAGTGCTCGCCTTGCACAATCGAATACCTTTA
27430 27440 27450 27460 27470 27480
ATCATCAACTTTGCGGCGAGTTAAGAACACTACGTCAGCATTGGGGCATGGATGATGTAG
27490 27500 27510 27520 27530 27540
CTAACGGTACGTTTCGAGCAGATCTGGGAAACCTATAACCATTAAAGGCAACGATTGAACAGG
27550 27560 27570 27580 27590 27600
TCGAAATTGCCTCTGCAGATGAAGATCGTGTGATTGCACCATTATCAATACACCTGATA
27610 27620 27630 27640 27650 27660
GCTTGTGTAGCCGGTTATCCAGAAGCCTGTCAGCGAGTCATTAAAGAATTTAGGTGTGC
27670 27680 27690 27700 27710 27720
GTGCAATGGCATTGAATATGGCGAACGCAATTCACAGCGCGCCAGCTTATGCCGAATACG
27730 27740 27750 27760 27770 27780
ATCATATGGTTGAGCTATACCATATGGATGTTACTCCACGTATTAATACCAAGATGTATT
27790 27800 27810 27820 27830 27840
CAAGCTCATGTTATTTACCGATTCCACAACGCAGCAAAGCGATTTCCACAGTATTGCTA
27850 27860 27870 27880 27890 27900
AATGTTTGTGTGATGTGGTGGATTTCACGTTTGGTTAATACCTTACATGACAAAGGTG
27910 27920 27930 27940 27950 27960
CGCGGGTATTTCATTGAAATGGGTCCAGGTTCGTTTATGTAGCTGGGTAGATAAGATCT
27970 27980 27990 28000 28010 28020
TAGTTAATGGCGATGGCGATAATAAAAAGCAAAGCCAACATGTATCTGTTCTGTGAATG
28030 28040 28050 28060 28070 28080
CCAAAGGCACCAAGTGATGAACCTTACTTATATTCGTGCGATTGCTAAGTTAATTAGTCATG
28090 28100 28110 28120 28130 28140
GCGTGAATTTGAATTTAGATAGCTTGTTTAACGGGTCAATCCTGGTTAAAGCAGGCCATA
28150 28160 28170 28180 28190 28200
TAGCAAACACGAACAAATAGTCAACATCGATATCTAGCGCTGGTGAGTTATACCTCATT
28210 28220 28230 28240 28250 28260
GTTGAAATATGGATTTAAAGAGAGTAATTATGGAAAATATTGCAGTAGTAGGTATTGCTA
28270 28280 28290 28300 28310 28320
ATTTGTTCCCGGCTCACAAGCACCGGATCAATTTTGGCAGCAATTGCTTGAACAACAAG
28330 28340 28350 28360 28370 28380
ATTGCCCGAGTAAGGCGACCGCTGTTCAAATGGGCGTTGATCCTGCTAAATATACCGCCA
28390 28400 28410 28420 28430 28440
ACAAAGGTGACACAGATAAATTTTACTGTGTGCACGGCGTTACATCAGTGATTTCATT
28450 28460 28470 28480 28490 28500
TTGATGCTTCAGGTTATCAACTCGATAATGATTATTTAGCCGGTTTAGATGACCTTAATC

Fig. 5

28510 28520 28530 28540 28550 28560
AATGGGGGCTTTATGTTACGAAACAAGCCCTTACCGATGCGGGTTATTGGGGCAGTACTG

28570 28580 28590 28600 28610 28620
CACTAGAAAACGTGTGGTGTGATTTTAGGTAATTTGTCATTCCCAACTAAATCATCTAATC

28630 28640 28650 28660 28670 28680
AGCTGTTTATGCCTTTGTATCATCAAGTTGTTGATAATGCCTTAAAGGCGGTATTACATC

28690 28700 28710 28720 28730 28740
CTGATTTTCAATTAACGCATTACACAGCACCGAAAAAACACATGCTGACAATGCATTAG

28750 28760 28770 28780 28790 28800
TAGCAGGTTATCCAGCTGCATTGATCGCGCAAGCGGCGGGTCTTGGTGGTTCACATTTTG

28810 28820 28830 28840 28850 28860
CACTGGATGCGGCTTGTGCTTCATCTTGTATAGCGTTAAGTTAGCGTGTGATTACCTGC

28870 28880 28890 28900 28910 28920
ATACGGGTAAAGCCAACATGATGCTTGCTGGTGCAGTATCTGCAGCAGATCCTATGTTTCG

28930 28940 28950 28960 28970 28980
TAAATATGGGTTTCTCGATATTCCAAGCTTACCCAGCTAACAATGTACATGCCCCGTTTG

28990 29000 29010 29020 29030 29040
ACCAAAATTACAAAGTCTATTTGCCGGTGAAGGCGCGGCATGATGGTATTGAAACGTC

29050 29060 29070 29080 29090 29100
AAAGTGATGCAGTACGTGATGGTGATCATATTTACGCCATTATTAAAGGCGGCGCATTAT

29110 29120 29130 29140 29150 29160
CGAATGACGGTAAAGGCGAGTTTGTATTAAAGCCCGAACACCAAGGGCCAAGTATTAGTAT

29170 29180 29190 29200 29210 29220
ATGAACGTGCTTATGCCGATGCAGATGTTGACCCGAGTACAGTTGACTATATTGAATGTC

29230 29240 29250 29260 29270 29280
ATGCAACGGGCACACCTAAGGGTGACAATGTTGAATTGCGTTCGATGGAAACCTTTTCA

29290 29300 29310 29320 29330 29340
GTCGCGTAAATAACAAACCATTACTGGGCTCGGTTAAATCTAACCTTGGTCATTTGTTAA

29350 29360 29370 29380 29390 29400
CTGCCGCTGGTATGCCTGGCATGACCAAAGCTATGTTAGCGCTAGGTAAAGGTCTTATTC

29410 29420 29430 29440 29450 29460
CTGCAACGATTAACTTAAAGCAACCACTGCAATCTAAAAACGGTACTTTACTGGCGAGC

29470 29480 29490 29500 29510 29520
AAATGCCAACGACGACTGTGTCTTGGCCAACAACCTCCGGGTGCCAAGGCAGATAAACCGC

29530 29540 29550 29560 29570 29580
GTACCGCAGGTGTGAGCGTATTTGGTTTTGGTGGCAGCAACGCCCATTTGGTATTACAAC

29590 29600 29610 29620 29630 29640
AGCCAACGCAAACTCGAGACTAATTTTAGTGTGCTAAACCACGTGAGCCTTTGGCTA

Fig. 5

29650 29660 29670 29680 29690 29700
TTATTGGTATGGACAGCCATTTTGGTAGTGCCAGTAATTTAGCGCAGTTCAAAACCTTAT

29710 29720 29730 29740 29750 29760
TAAATAATAATCAAAATACCTTCCGTGAATTACCAGAACAACGCTGGAAAGGCATGGAAA

29770 29780 29790 29800 29810 29820
GTAACGCTAACGTCATGCAGTCGTTACAATTACGCAAAGCGCCTAAAGGCAGTTACGTTG

29830 29840 29850 29860 29870 29880
AACAGCTAGATATTGATTTCTTGCGTTTTAAAGTACCGCCTAATGAAAAAGATTGCTTGA

29890 29900 29910 29920 29930 29940
TCCCGCAACAGTTAATGATGATGCAAGTGGCAGACAATGCTGCGAAAGACGGAGGTCTAG

29950 29960 29970 29980 29990 30000
TTGAAGGTTCGTAATGTTGCGGTATTAGTAGCGATGGGCATGGAAGTGAATTACATCAGT

30010 30020 30030 30040 30050 30060
ATCGTGGTTCGCGTTAATCTAACCACCCAAATTGAAGACAGCTTATTACAGCAAGGTATTA

30070 30080 30090 30100 30110 30120
ACCTGACTGTTGAGCAACGTGAAGAACTGACCAATATTGCTAAAGACGGTGTTCCTCGG

30130 30140 30150 30160 30170 30180
CTGCACAGCTAAATCAGTATACGAGTTTCATTGGTAATATTATGGCGTCACGTATTTTCGG

30190 30200 30210 30220 30230 30240
CGTTATGGGATTTTTCTGGTCCTGCTATTACCGTATCGGCTGAAGAAAACCTCTGTTTATC

30250 30260 30270 30280 30290 30300
GTTGTGTTGAATTAGCTGAAAATCTATTTCAAACCAGTGATGTTGAAGCCGTTATTATTG

30310 30320 30330 30340 30350 30360
CTGCTGTTGATTTGTCTGGTTCAATTGAAAACATTACTTTACGTCAGCACTACGGTCCAG

30370 30380 30390 30400 30410 30420
TTAATGAAAAGGGATCTGTAAGTGAATGTGGTCCGGTTAATGAAAGCAGTTCAGTAACCA

30430 30440 30450 30460 30470 30480
ACAAATATTCTTGATCAGCAACAATGGCTGGTGGGTGAAGGCGCAGCGGTATTGTCGTTA

30490 30500 30510 30520 30530 30540
AACCGTCATCGCAAGTCACTGCTGAGCAAGTTTATGCGCGTATTGATGCGGTGAGTTTTG

30550 30560 30570 30580 30590 30600
CCCCTGGTAGCAATGCGAAAGCAATTACGATTGCAGCGGATAAAGCATTAACACTTGCTG

30610 30620 30630 30640 30650 30660
GTATCAGTGCTGCTGATGTAGCTAGTGTTGAAGCACATGCAAGTGGTTTTAGTGCCGAAA

30670 30680 30690 30700 30710 30720
ATAATGCTGAAAAAACCGGTTACCGACTTTATACCCAAGCGCAAGTATCAGTTCGGTGA

30730 30740 30750 30760 30770 30780
AAGCCAATATTGGTCATACGTTTAAATGCCTCQGGTATGGCGAGTATTATTAACCGGCGC

Fig. 5

30790 30800 30810 30820 30830 30840
TGCTGTTAGATCAGAATACGAGTCAAGATCAGAAAAGCAAACATATTGCTATTAACGGTC

30850 30860 30870 30880 30890 30900
TAGGTCGTGATAACAGCTGCGCGCATCTTATCTTATCGAGTTCAGCGCAAGCGCATCAAG

30910 30920 30930 30940 30950 30960
TTGCACCAGCGCCTGTATCTGGTATGGCCAAGCAACGCCCCACAGTTAGTTAAAACCATCA

30970 30980 30990 31000 31010 31020
AACTCGGTGGTCAGTTAATTAGCAACGCGATTGTTAACAGTGCAGTTTCATCTTTACACG

31030 31040 31050 31060 31070 31080
CTATTAAAGCGCAGTTTGCCGGTAAGCACTTAAACAAAGTTAACCAGCCAGTGATGATGG

31090 31100 31110 31120 31130 31140
ATAACCTGAAGCCCCAAGGTATTAGCGCTCATGCAACCAATGAGTATGTGGTGACTGGAG

31150 31160 31170 31180 31190 31200
CTGCTAACACTCAAGCTTCTAACATTCAAGCATCTCATGTTCAAGCGTCAAGTCATGCAC

31210 31220 31230 31240 31250 31260
AAGAGATAGCACCAAACCAAGTTCAAAATATGCAAGCTACAGCAGCCGCTGTAAGTTCAC

31270 31280 31290 31300 31310 31320
CCCTTTCTCAACATCAACACACAGCGCAGCCGCTAGCGGCACCGAGCGTTGTTGGAGTGA

31330 31340 31350 31360 31370 31380
CTGTGAAACATAAAGCAAGTAACCAAATTCATCAGCAAGCGTCTACGCATAAAGCATTTT

31390 31400 31410 31420 31430 31440
TAGAAAGTCGTTTAGCTGCACAGAAAACCTATCGCAACTTGTTGAATTGCAAACCAAGC

31450 31460 31470 31480 31490 31500
TGTCATCCAAACTGGTAGTGACAATACATCTAACAATACTGCGTCAACAAGCAATACAG

31510 31520 31530 31540 31550 31560
TGCTAACAAATCCTGTATCAGCAACGCCATTAACTTGTGTCTAATGCGCCTGTAGTAG

31570 31580 31590 31600 31610 31620
CGACAAACCTAACCAGTACAGAAGCAAAAGCGCAAGCAGCTGCTACACAAGCTGGTTTTTC

31630 31640 31650 31660 31670 31680
AGATAAAAGGACCTGTTGGTTACAACCTATCCACCGCTGCAGTTAATTGAACGTTATAATA

31690 31700 31710 31720 31730 31740
AACCAGAAAACGTGATTTACGATCAAGCTGATTTGGTTGAATTGCTGAAGGTGATATTG

31750 31760 31770 31780 31790 31800
GTAAGGTATTTGGTGTGAATACAATATTATTGATGGCTATTGCGCTCGTGTACGTCTGC

31810 31820 31830 31840 31850 31860
CAACCTCAGATTACTTGTAGTAACACGTGTTACTGAACTTGATGCCAAGGTGCATGAAT

31870 31880 31890 31900 31910 31920
ACAAGAAATCATACATGTGTACTGAATATGATGCGCTGTTGATGCACCGTTCTTAATTG

Fig. 5

31930 31940 31950 31960 31970 31980
ATGGTCAGATCCCCTTGGTCTGTTGCCGTCGAATCAGGCCAGTGTGATTTGATGTTGATTT

31990 32000 32010 32020 32030 32040
CATATATCGGTATTGATTTCCAAGCGAAAGGCGAACGTGTTTACCGTTTACTTGATTGTG

32050 32060 32070 32080 32090 32100
AATTAACTTTCCTTGAAGAGATGGCTTTTGGTGGCGATACTTTACGTTACGAGATCCACA

32110 32120 32130 32140 32150 32160
TTGATTTCGTATGCACGTAACGGCGAGCAATTATTATTCTTCTTCCATTACGATTGTTACG

32170 32180 32190 32200 32210 32220
TAGGGGATAAGAAGGTACTTATCATGCGTAATGGTTGTGCTGGTTTCTTTACTGACGAAG

32230 32240 32250 32260 32270 32280
AACTTTCTGATGGTAAAGGCGTTATTACATAACGACAAAGACAAAGCTGAGTTTAGCAATG

32290 32300 32310 32320 32330 32340
CTGTTAAATCATCATTACGCCGTTATTACAACATAACCGTGGTCAATACGATTATAACG

32350 32360 32370 32380 32390 32400
ACATGATGAAGTTGGTTAATGGTGATGTTGCCAGTTGTTTTGGTCCGCAATATGATCAAG

32410 32420 32430 32440 32450 32460
GTGGCCGTAATCCATCATTGAAATTCTCGTCTGAGAAGTTCTTGATGATTGAACGTATTA

32470 32480 32490 32500 32510 32520
CCAAGATAGACCCAACCGGTGGTTCATTGGGGACTAGGCCTGTTAGAAGGTCAGAAAGATT

32530 32540 32550 32560 32570 32580
TAGACCCTGAGCATTGGTATTTCCCTTGTCACTTTAAAGGTGATCAAGTAATGGCTGGTT

32590 32600 32610 32620 32630 32640
CGTTGATGTCGGAAGGTTGTGGCCAAATGGCGATGTTCTTCATGCTGTCTCTTGGTATGC

32650 32660 32670 32680 32690 32700
ATACCAATGTGAACAACGCTCGTTTCCAACCACTACCAGGTGAATCACAAACGGTACGTT

32710 32720 32730 32740 32750 32760
GTCGTGGGCAAGTACTGCCACAGCGCAATACCTTAACTTACCGTATGGAAGTTACTGCGA

32770 32780 32790 32800 32810 32820
TGGGTATGCATCCACAGCCATTTCATGAAAGCTAATATTGATATTTTGCTTGACGGTAAAG

32830 32840 32850 32860 32870 32880
TGGTTGTTGATTTCAAAAACCTTGAGCGTGATGATCAGCGAACAAGATGAGCATTACGATT

32890 32900 32910 32920 32930 32940
ACCCTGTAACACTGCCGAGTAATGTGGCGCTTAAAGCGATTACTGCACCTGTTGCGTCAG

32950 32960 32970 32980 32990 33000
TAGCACCAGCATCTTCACCCGCTAACAGCGCGGATCTAGACGAACGTGGTGTGAACCGT

33010 33020 33030 33040 33050 33060
TTAAGTTTCCTGAACGTCCGTTAATGCGTGTTGAGTCAGACTTGCTGCACCGAAAAGCA

Fig. 5

33070 33080 33090 33100 33110 33120
AAGGTGTGACACCGATTAAGCATTTTGAAGCGCCTGCTGTTGCTGGTCATCATAGAGTGC

33130 33140 33150 33160 33170 33180
CTAACCAAGCACCGTTTACACCTTGGCATATGTTTGAGTTTGCACGGGTAATATTTCTA

33190 33200 33210 33220 33230 33240
ACTGTTTCGGTCTCTGATTTTGATGTTTATGAAGTTCGTATTCCACCTCGTACACCTTGTG

33250 33260 33270 33280 33290 33300
GCGATTTACAAGTTGTTACTCAGGTTGTAGAAGTGCAGGGCGAACGTCTTGATCTTAAAA

33310 33320 33330 33340 33350 33360
ATCCATCAAGCTGTGTAGCTGAATACTATGTACCGGAAGACGCTTGGTACTTTACTAAAA

33370 33380 33390 33400 33410 33420
ACAGCCATGAAAACCTGGATGCCTTATTCATTAATCATGGAAATTGCATTGCAACCAAATG

33430 33440 33450 33460 33470 33480
GCTTTATTTCTGGTTACATGGGCACGACGCTTAAATACCCTGAAAAAGATCTGTTCTTCC

33490 33500 33510 33520 33530 33540
GTAACCTTGATGGTAGCGGCACGTTATTAAAGCAGATTGATTTACGCGGCAAGACCATTG

33550 33560 33570 33580 33590 33600
TGAATAAATCAGTCTTGGTTAGTACGGCTATTGCTGGTGGCGGATTATTCAAAGTTTCA

33610 33620 33630 33640 33650 33660
CGTTTGATATGTCTGTAGATGGCGAGCTATTTTATACTGGTAAAGCTGTATTTGGTTACT

33670 33680 33690 33700 33710 33720
TTAGTGGTGAATCACTGACTAACCAACTGGGCATTGATAACGGTAAAACGACTAATGCGT

33730 33740 33750 33760 33770 33780
GGTTTGTTGATAACAATACCCCCGACGGAATATTGATGTGTTTGATTAACTAATCAGT

33790 33800 33810 33820 33830 33840
CATTGGCTCTGTATAAAGCGCCTGTGGATAAACCGCATTATAAATTGGCTGGTGGTCAGA

33850 33860 33870 33880 33890 33900
TGAACCTTATCGATACAGTGTGAGTGGTTGAAGCGGTGGTAAAGCGGGCGTGGCTTATG

33910 33920 33930 33940 33950 33960
TTTATGGCGAACGTACGATTGATGCTGATGATTGGTTCTTCCGTTATCACTTCCACCAAG

33970 33980 33990 34000 34010 34020
ATCCGGTGATGCCAGGTTTCATTAGGTGTTGAAGCTATTATTGAGTTGATGCAGACCTATG

34030 34040 34050 34060 34070 34080
CGCTTAAAAATGATTTGGGTGGCAAGTTTGCTAACCACGTTTCATTGCGCCGATGACGC

34090 34100 34110 34120 34130 34140
AAGTTGATTGGAAATACCGTGGGCAAATTACGCCGCTGAATAAACAGATGTCACTGGACG

34150 34160 34170 34180 34190 34200
TGCATATCACTGAGATCGTGAATGACGCTGGTGAAGTGCGAATCGTTGGTGATGCGAATC

Fig. 5

34210 34220 34230 34240 34250 34260
TGCTAAAGATGGTCTGCGTATTTATGAAGTTAAAACATCGTTTTAAGTATTGTTGAAG

34270 34280 34290 34300 34310 34320
CGTAAAGGGTCAAGTGTAACGTGCTTAAGCGCCGCATTGGTTAAAGACGCTTTGCACGCC

34330 34340 34350 34360 34370 34380
GTGAATCCGTCCATGGAGGCTTGGGGTTGGCATCCATGCCAACACAGCAAGCTTACTTT

34390 34400 34410 34420 34430 34440
AATCAATACGGCTTGGTGTCCATTTAGACGCCCTCGAACTTAGTAGTTAATAGACAAAATA

34450 34460 34470 34480 34490 34500
ATTTAGCTGTGGAATGAATATAGTAAGTAATCATTCGGCAGCTACAAAAAAGGAATTAAG

34510 34520 34530 34540 34550 34560
AATGTCGAGTTTAGGTTTTACAATAACAACGCAATTAAGTGGGCTTGAAAGTAGATCC

34570 34580 34590 34600 34610 34620
AGCGTCAGTTCATACACAAGATGCAGAAATTAAAGCAGCTTAAATGGATCTAACTAAACC

34630 34640 34650 34660 34670 34680
TCTCTATGTGGCGAATAATTCAGGCGTAACTGGTATAGCTAATCATACGTCAGTAGCAGG

34690 34700 34710 34720 34730 34740
TGCGATCAGCAATAACATCGATGTTGATGTATTGGCGTTTGCGCAAAAGTTAAACCCAGA

34750 34760 34770 34780 34790 34800
AGATCTGGGGTGATGATGCTTACAAGAAACAGCACGGCGTTAAATATGCTTATCATGGCGG

34810 34820 34830 34840 34850 34860
TGCGATGGCAAATGGTATTGCCTCGGTTGAATTGGTTGTTGCGTTAGGTAAAGCAGGGCT

34870 34880 34890 34900 34910 34920
GTTATGTTCAATTTGGTGCTGCAGGTCTAGTGCCTGATGCGGTTGAAGATGCAATTCGTCG

34930 34940 34950 34960 34970 34980
TATTCAAGCTGAATTACCAAATGGCCCTTATGCGGTTAACTTGATCCATGCACCAGCAGA

34990 35000 35010 35020 35030 35040
AGAAGCATTAGAGCGTGGCGCGGTTGAACGTTTCCTAAAACCTGGCGTCAAGACGGTAGA

35050 35060 35070 35080 35090 35100
GGCTTCAGCTTACCTTGTTTAACTGAACACATTGTTTGGTATCGTGCTGCTGGTCTAAC

35110 35120 35130 35140 35150 35160
TAAAAACGCAGATGGCAGTGTTAATATCGGTAACAAGGTTATCGCTAAAGTATCGCGTAC

35170 35180 35190 35200 35210 35220
CGAAGTTGGTCGCCGCTTTATGGAACCTGCACCGCAAAAATTACTGGATAAGTTATTAGA

35230 35240 35250 35260 35270 35280
ACAAAATAAGATCACCCCTGAACAAGCTGCTTTAGCGTTGCTTGTACCTATGGCTGATGA

35290 35300 35310 35320 35330 35340
TATTACTGGGGAAGCGGATTCTGGTGGTCATACAGATAACCGTCCGTTTTTAACATTATT

Fig. 5

35350 35360 35370 35380 35390 35400
ACCGACGATTATTGGTCTGCGTGATGAAGTGCAAGCGAAGTATAACTTCTCTCCTGCATT

35410 35420 35430 35440 35450 35460
ACGTGTTGGTGCTGGTGGTGGTATCGGAACGCCTGAAGCAGCACTCGCTGCATTTAACAT

35470 35480 35490 35500 35510 35520
GGGCGCGCTTATATCGTTCTGGGTCTGTGAATCAGGCGTGTGTTGAAGCGGGTGCATC

35530 35540 35550 35560 35570 35580
TGAATATACTCGTAACTGTTATCGACAGTTGAAATGGCTGATGTGACTATGGCACCTGC

35590 35600 35610 35620 35630 35640
TGCAGATATGTTTGAATGGGTGTGAAGCTGCAAGTATTAACGCGGTTCTATGTTTCGC

35650 35660 35670 35680 35690 35700
GATGCGTGCGAAGAACTGTATGACTTGTATGTGGCTTATGACTCGATTGAAGATATCCC

35710 35720 35730 35740 35750 35760
AGCTGCTGAACGTGAGAAGATTGAAAAACAAATCTTCCGTGCAAACCTAGACGAGATTTG

35770 35780 35790 35800 35810 35820
GGATGGCACTATCGCTTTCTTTACTGAACGCGATCCAGAAATGCTAGCCCGTGCAACGAG

35830 35840 35850 35860 35870 35880
TAGTCCTAAACGTAAATGGCACTTATCTTCCGTTGGTATCTTGGCCTTTCTTCACGCTG

35890 35900 35910 35920 35930 35940
GTCAAACACAGGCGAGAAGGGACGTGAAATGGATTATCAGATTGCGGCAGGCCCAAGTTT

35950 35960 35970 35980 35990 36000
AGGTGCATTCAACAGCTGGGTGAAAGGTTCTTACCTTGAAGACTATACCCGCCGTGGCGC

36010 36020 36030 36040 36050 36060
TG TAGATGTTGCTTTGCATATGCTTAAAGGTGCTGCGTATTTACAACGTGTAAACCAAGTT

36070 36080 36090 36100 36110 36120
GAAATTGCAAGGTGTTAGCTTAAGTACAGAATTGGCAAGTTATCGTACGAGTGATTAATG

36130 36140 36150 36160 36170 36180
TTACTTGATGATATGTGAATTAATTAAAGCGCCTGAGGGCGCTTTTTTTGGTTTTTAACT

36190 36200 36210 36220 36230 36240
CAGGTGTTGTAACGCGAAATTGCCCCCTTCAAGTTAGATCGATTACTCACTCACAATATG

36250 36260 36270 36280 36290 36300
TTGATATCGCACTTGCCATATACTTGCTCATCCAAAGCCCTATATTGATAATGGTGTTAA

36310 36320 36330 36340 36350 36360
TAGTCTTTAATATCCGAGTCTTTCTTCAGCATAATACTAATATAGAGACTCGACCAATGT

36370 36380 36390 36400 36410 36420
TAAACACAACAAAGAATATATTCTTGTGTACTGCCTTATTATTAACGAGTGCGAGTACGA

36430 36440 36450 36460 36470 36480
CAGCTACTACGCTAAACAATTCGATATCAGCAATTGAACAACGTATTTCTGGTCGTATCG

Fig. 5

36490 36500 36510 36520 36530 36540
GTGTGGCTGTTTTAGATACGCAAAATAAACAAACGTGGGCTTACAATGGTGATGCACATT

36550 36560 36570 36580 36590 36600
TTCCGATGATGAGTACATTCAAAACCCCTCGCTTGCGCGAAAATGCTAAGTGAATCGACAA

36610 36620 36630 36640 36650 36660
ATGGTAATCTGGATCCCAGTACTAGCTCATTGATAAAGGCTGAAGAATTAATCCCTTGGT

36670 36680 36690 36700 36710 36720
CACCAGTCACTAAAACGTTTGTGAATAACACTATTACAGTGGCGAAAGCGTGTGAAGCAA

36730 36740 36750 36760 36770 36780
CAATGCTGACCAGTGATAATACGCGGCTAATATTGTTTTACAGTATATCGGAGGCCCTC

36790 36800 36810 36820 36830 36840
AAGGCGTTACTGCATTCTTGCGAGAAATTGGTGATGAAGAGAGTCAGTTAGATCGTATAG

36850 36860 36870 36880 36890 36900
AACCTGAATTGAATGAAGCTAAGGTCGGAGACTTGCGTGATACCACGACACCGAAAGCCA

36910 36920 36930 36940 36950 36960
TAGTTACCACGCTCAACAACTACTACTTGGTGATGTTCTACTTGATTTGGATAAAAACC

36970 36980 36990 37000 37010 37020
AACTTAAACATGGATGCAAAATAATAAAGTGTGATCCTTTACTGCGTTCTATATTAC

37030 37040 37050 37060 37070 37080
CGCAAGGCTGGTTTATTGCCGACCGCTCAGGTGCGGGTGGTAATGGTTCTCGAGGTATAA

37090 37100 37110 37120 37130 37140
CTGCTATGCTTTGGCACTCCGAGCGTCAACCGCTAATCATCAGTATTTATTTAACCAGAA

37150 37160 37170 37180 37190 37200
CTGAGTTAGCAATGCAATGCGCAATGAGATTATTGTTGAGATCGGTAAGCTGATATTCA

37210 37220 37230 37240 37250 37260
AAGAATACGCGGTGAAATAATAAGTTATTTTTTGATAATACTTTAACGAGCGTAGCTATC

37270 37280 37290 37300 37310 37320
GAAGTGAGGGCGTCAATTAGACACCTTTGCTTCCCCTACAAAATCTAATGTGTATTACCT

37330 37340 37350 37360 37370 37380
CGGCTAGTACAATTGCCCTAAGTTATTTCTGTCCAGCTTTGGCTTAGTGCAATTGCGTTA

37390 37400 37410 37420 37430 37440
GCCAATGTGAACACCAAGGGACTTTGTGTCGTACCATAACTACCAAGCGACTTTGTCGTTTT

37450 37460 37470 37480 37490 37500
TATCTTTTCTTAGACAAACAGAGGTTAAATGAGTGACGCCTTCCAAATCACAGGAATGAA

37510 37520 37530 37540 37550 37560
TCCGATTTCAATAAAATCTAACCCGTACCAACTCCGTACAAGTTGATCTTTAGTTGTTT

37570 37580 37590 37600 37610 37620
AAAATCTATAATAAATTCAATTACGGAATTAATCCGTACAACCTGGAGGTTTTATGGCTAC

Fig. 5

37630 37640 37650 37660 37670 37680
TGCAAGACTTGATATCCGTTTGGATGAAGAAATCAAAGCTAAGGCTGAGAAAGCATCAGC

37690 37700 37710 37720 37730 37740
TTTACTCGGCTTAAAAAGTTTAACCGAATACGTTGTTTCGCTTAATGGACGAAGATTCAAC

37750 37760 37770 37780 37790 37800
TAAAGTAGTTTCTGAGCATGAGAGTATTACCGTTGAAGCGAATGTATTGACCAATTTAT

37810 37820 37830 37840 37850 37860
GGCTGCTTGTGATGAAGCGAAAGCCCCAAATAAAGCATTACTTGAAGCCGCTGTATTTAC

37870 37880 37890 37900 37910 37920
TCAGAATGGTGAGTTTAAGTGAGTTATTCCAAACGTTTCAAAGAAGTGGATAAATCAAAA

37930 37940 37950 37960 37970 37980
CATGACAGAGCATCATTTGACTGTGGCGAAAAAGAGCTAAATGATTTTATCCAAACTCAA

37990 38000 38010 38020 38030 38040
GCAGCCAAACATATGCAAGCAGGTATTAGCCGCACTCTGGTTTTACCTGCTTCTGCGCCG

38050 38060 38070 38080 38090 38100
TTACCAAACAAAAATATCCAATTTGCTCATTTTATAGTATCGCGCCAAGCTCAATTAGC

38110 38120 38130 38140 38150 38160
CGCGATACGTTACCACAAGCAATGGCTAAAAAGTTACCACGTTATCCTATCCCTGTTTTT

38170 38180 38190 38200 38210 38220
CTTTTGGCTCAACTTGCCGTCCATAAAGAGTTTCATGGGAGTGGGTTAGGCAAAGTTAGC

38230 38240 38250 38260 38270 38280
TTAATTAAAGCGTTAGAGTACCTTTGGGAAATTAAGTCTCACATGAGAGCTTACGCCATC

38290 38300 38310 38320 38330 38340
GTTGTTGATTGTTTAACTGAACAAGCTGAGTCATTCTACGCTAAATATGGTTTCGACGTT

38350 38360 38370 38380 38390 38400
CTCTGCGAAATAAATGGTCGAGTAAGAATGTTTCATATCAATGAAAACAGTCAATCAGTTA

38410 38420 38430 38440 38450 38460
TTCACTTAACAGTAAGAGTTAGTATAACAGTTGTATGAATTAAATTTATTATATTTCGGTA

38470 38480 38490 38500 38510 38520
ATCTCATTGCGATCACGCTAGAAGTGCAGCGGGTCAGACCGAGGCCACAATAGCAGCCG

38530 38540 38550 38560 38570 38580
TTACGTTTAGGGGATGACTTAAAAAGATAACTACTACGTCAGTGGCGATCCTAGAGGATT

38590 38600 38610 38620 38630 38640
AAAGGTTTATGATTCAACAACATTTATTTATTGTGCTTAATTTTTTCTATCCAATATGCGC

38650 38660 38670 38680 38690 38700
AAGCTGTAAATATCACTGAAGTAGACTTTTATGTCAGTGATGATATCCCTAAAGATGTTG

38710 38720 38730 38740 38750 38760
CCAAATTAAAGATAGGTGAATCCATAACGAACTCCAGCCTTATTCTAAGTAACTCATCTA

Fig. 5

38770 38780 38790 38800 38810 38820
TTCCACTCTCGCGGGAGACGGGTAACATATATTACTCTTCATCAATTGCTAACTTGAAC

38830 38840 38850 38860 38870 38880
ATGACTCGATAGAATTTGTTATGGCTCAATTGATGGCCGAAGATTCCAGCCTTTACAAGA

38890 38900 38910 38920 38930 38940
TGCTGGTAAATAGCGATAGGTTGTCCGTGCTAGTAATGACATCTTCCCAGTCCACAGATC

38950 38960 38970 38980 38990 39000
TCTATGGCTCGACTTACTCGGCTTATTTTCTAATGTTGCGGTCATCGATTTGAATTGTG

39010 39020 39030 39040 39050 39060
ACTCGCTAACTTTAGAACATGAGCTCGGCCATCTATACGGAGCTGAACATGAAGAAATAT

39070 39080 39090 39100 39110 39120
ATGACGACTATGTCTTCTATGCTGCGATATGTGGAGACTATACGACTATCATGAACCTCTA

39130 39140 39150 39160 39170 39180
TGCAGCCTGAAATGAAAGAAAAACAAATGATAAAGGCATATTCATTCCCTGAATTAAAAG

39190 39200 39210 39220 39230 39240
TGGATGGCTTGCGAGTGCAGGAAATGAAATACGAATAACAAAAAGGTTATTTTAGACAATA

39250 39260 39270 39280 39290 39300
TTGGTCGGTTTAGATAGGATTGGGATATTATTCTCATTTCGGCTCTACTTAGTGCTGTTAT

39310 39320 39330 39340 39350 39360
TATGAGTGCCAGTGCTTCTATCTACGATATTGGTCTTAACAAGTATTTATCTATAGACGC

39370 39380 39390 39400 39410 39420
TAAGGTGTTATGTATTTAAGGGATGTTCAAGATGAACTAGGTGTAAACGATGTATAGTT

39430 39440 39450 39460 39470 39480
GTATAACATTTTTTCAACGGTTGGAACGTTGATTCTATCGGGTAACAAGACCGCGACGA

39490 39500 39510 39520 39530 39540
TCCGCGATAAGTCCGATAGTCATTACTTAGTTGGTCAGATGTTAGATGCTTGTACTCACG

39550 39560 39570 39580 39590 39600
AAGATAATCGGAAAATGTGTCAAATAGAAATACTGAGCATTGAATATGTGACGTTTAGTG

39610 39620 39630 39640 39650 39660
AATTAAACCGTGCGCACGCCAATGCTGAAGGTTTACCGTTTTTGTGTTATGCTTAAGTGGA

39670 39680 39690 39700 39710 39720
TAGTTCGAAAGATTTATCCGACTTCAAATGATTTATTTTTCATAAGTTTCAGAGTTGTAA

39730 39740 39750 39760 39770 39780
CTATCGATATCTTATAAGTCTTAGTGCACAAAACAGAACTATTTATAGCGCTCAAGAAGG

39790 39800 39810 39820 39830 39840
CGATAATTTGATAATGAATTATCGCCTTGTTACTATTAAGAGACTTTAAATGACTGAGAT

39850 39860 39870 39880 39890 39900
ATAAGATATGACACGGAAGACATATTGATCACAGGCGCAAGTTCAGGGTTGGGCCGAGG

Fig. 5

39910	39920	39930	39940	39950	39960
TATGGCCATCGAATTTGCAAAATCAGGTCATAACTTAGCACTTTGTGCACGTAGACTTGA					
39970	39980	39990	40000	40010	40020
TAATTTAGTTGCACTGAAAGCAGAACTCTTAGCCCTCAATCCTCACATCCAAATCGAAAT					
40030	40040	40050	40060	40070	40080
AAAACCTCTTGATGTCAATGAACATGAACAAGTCTTCACTGTTTTCCATGAATTCAAAGC					
40090	40100	40110	40120	40130	
TGAATTTGGTACGCTTGATCGTATTATTGTTAATGCTGGATTAGGCAAGGGTGGATCC					

Fig. 5

10 20 30 40 50 60
AAATGCAATTAATTATGGCGTAAATAGAGTGAAAACATGGCTAATATTCACCTAAGTCCTG

70 80 90 100 110 120
AATTTTATATAAAGTTTAACTCTGTTATTTTAGCGTTTACCTGGTCTTATCAGTGAGGTTT

130 140 150 160 170 180
ATAGCCATTATTAGTGGGATTGAAGTGATTTTAAAGCTATGTATATTATTGCAAATATA

190 200 210 220 230 240
AATTGTAACAATTAAGACTTTGGACACTTGAGTTCAATTTCGAATTGATTGGCATAAAAT

250 260 270 280 290 300
TTAAAACAGCTAAATCTACCTCAATCATTTTAGCAAATGTATGCAGGTAGATTTTTTTTCG

310 320 330 340 350 360
CCATTTAAGAGTACACTTGTACGCTAGGTTTTTGTGTTAGTGTGCAAATGAACGTTTTTGAT

370 380 390 400 410 420
GAGCATTGTTTTTAGAGCACAAAATAGATCCTTACAGGAGCAATAACGCAATGGCTAAAA

430 440 450 460 470 480
AGAACACCACATCGATTAAAGCACGCCAAGGATGTGTTAAGTAGTGATGATCAACAGTTAA

490 500 510 520 530 540
ATTCTCGCTTGCAAGAATGTCCGATTGCCATCATTGGTATGGCATCGGTTTTTGCAGATG

550 560 570 580 590 600
CTAAAAACTTGGATCAATTCTGGGATAACATCGTTGACTCTGTGGACGCTATTATTGATG

610 620 630 640 650 660
TGCCTAGCGATCGCTGGAACATTGACGACCATTACTCGGCTGATAAAAAAGCAGCTGACA

670 680 690 700 710 720
AGACATACTGCAAACGCGGTGGTTTCATTCCAGAGCTTGATTTTGATCCGATGGAGTTTG

730 740 750 760 770 780
GTTTACCGCCAAATATCCTCGAGTTAACTGACATCGCTCAATTGTTGTCATTAATTGTTG

790 800 810 820 830 840
CTCGTGATGTATTAAGTGATGCTGGCATTGGTAGTGATTATGACCATGATAAAATTGGTA

850 860 870 880 890 900
TCACGCTGGGTGTCGGTGGTGGTCAGAAACAAATTCGCCATTAACGTCGCGCCTACAAG

910 920 930 940 950 960
GCCCCGTATTAGAAAAAGTATTAAGCCTCAGGCATTGATGAAGATGATCGCGCTATGA

970 980 990 1000 1010 1020
TCATCGACAAATTTAAAAAGCCTACATCGGCTGGGAAGAGAACTCATTCCCAGGCATGC

1030 1040 1050 1060 1070 1080
TAGGTAACGTTATTGCTGGTCGTATCGCCAATCGTTTTGATTTTGGTGGTACTAACTGTG

1090 1100 1110 1120 1130 1140
TGGTTGATGCGGCATGCGCTGGCTCCCTTGCACTGTAAAATGGCGATCTCAGACTTAC

Fig. 6

1150 1160 1170 1180 1190 1200
TTGAATATCGTTTCAGAAAGTCATGATATCGGGTGGTGTATGTTGTGATAACTCGCCATTCA
1210 1220 1230 1240 1250 1260
TGTATATGTCATTCTCGAAAACACCAGCATTACCACCAATGATGATATCCGTCCTGTTG
1270 1280 1290 1300 1310 1320
ATGACGATTCAAAAGGCATGCTGGTTGGTGAAGGTATTGGCATGATGGCGTTTAAACGTC
1330 1340 1350 1360 1370 1380
TTGAAGATGCTGAACGTGACGGCGACAAAATTTATTCTGTACTGAAAGGTATCGGTACAT
1390 1400 1410 1420 1430 1440
CTTCAGATGGTCGTTTCAAATCTATTTACGCTCCACGCCCAGATGGCCAAGCAAAAGCGC
1450 1460 1470 1480 1490 1500
TAAAACGTGCTTATGAAGATGCCGGTTTTGCCCCCTGAAACATGTGGTCTAATTGAAGGCC
1510 1520 1530 1540 1550 1560
ATGGTACGGGTACCAAAGCGGGTGATGCCGAGAAATTTGCTGGCTTGACCAAACACTTTG
1570 1580 1590 1600 1610 1620
GCGCCGCCAGTGATGAAAAGCAATATATCGCCTTAGGCTCAGTTAAATCGCAAATTGGTC
1630 1640 1650 1660 1670 1680
ATACTAAATCTGCGGCTGGCTCTGCGGGTATGATTAAGGCGGCATTAGCGCTGCATCATA
1690 1700 1710 1720 1730 1740
AAATCTTACCTGCAACGATCCATATCGATAAACCAAGTGAAGCCTTGATATCAAAAACA
1750 1760 1770 1780 1790 1800
GCCCCGTTATACCTAAACAGCGAAACGCGTCCTTGGATGCCACGTGAAGATGGTATTCCAC
1810 1820 1830 1840 1850 1860
GTCGTGCAGGTATCAGCTCATTTGGTTTTGGCGGCACCAACTTCCATATTATTTTAGAAG
1870 1880 1890 1900 1910 1920
AGTATCGCCCAGGTCACGATAGCGCATATCGCTTAAACTCAGTGAGCCAAACTGTGTTGA
1930 1940 1950 1960 1970 1980
TCTCGGCAAACGACCAACAAGGTATTGTTGCTGAGTTAAATAACTGGCGTACTAAACTGG
1990 2000 2010 2020 2030 2040
CTGTCGATGCTGATCATCAAGGGTTTGTATTTAATGAGTTAGTGACAACGTGGCCATTAA
2050 2060 2070 2080 2090 2100
AAACCCCATCCGTTAACCAAGCTCGTTTAGGTTTTGTTGCGCGTAATGCAAATGAAGCGA
2110 2120 2130 2140 2150 2160
TCGCGATGATTGATACGGCATTGAAACAATTCAATGCGAACGCAGATAAAATGACATGGT
2170 2180 2190 2200 2210 2220
CAGTACCTACCGGGGTTTACTATCGTCAAGCCGGTATTGATGCAACAGGTAAAGTGTTG
2230 2240 2250 2260 2270 2280
CGCTATTCTCAGGGCAAGGTTTCGCAATACGTGAACATGGGTCGTGAATTAACCTGTAAC

Fig. 6

2290 2300 2310 2320 2330 2340
TCCCAAGCATGATGCACAGTGCTGCGGCGATGGATAAAGAGTTCAGTGCCGCTGGTTTAG

2350 2360 2370 2380 2390 2400
GCCAGTTATCTGCAGTTACTTTCCTATCCCTGTTTATACGGATGCCGAGCGTAAGCTAC

2410 2420 2430 2440 2450 2460
AAGAAGAGCAATTACGTTTAAACGCAACATGCGCAACCAGCGATTGGTAGTTTGAGTGTG

2470 2480 2490 2500 2510 2520
GTCTGTTCAAAACGTTTAAAGCAAGCAGGTTTAAAGCTGATTTTGCTGCCGGTCATAGTT

2530 2540 2550 2560 2570 2580
TCGGTGAGTTAACCGCATTATGGGCTGCCGATGTATTGAGCGAAAGCGATTACATGATGT

2590 2600 2610 2620 2630 2640
TAGCGCGTAGTCGTGGTCAAGCAATGGCTGCGCCAGAGCAACAAGATTTTGATGCAGGTA

2650 2660 2670 2680 2690 2700
AGATGGCCGCTGTTGTTGGTGATCCAAAGCAAGTCGCTGTGATCATTGATACCCTTGATG

2710 2720 2730 2740 2750 2760
ATGTCTCTATTGCTAAGTTCAACTCGAATAACCAAGTTGTTATTGCTGGTACTACGGAGC

2770 2780 2790 2800 2810 2820
AGGTTGCTGTAGCGGTTACAACCTTAGGTAATGCTGGTTTCAAAGTTGTGCCACTGCCGG

2830 2840 2850 2860 2870 2880
TATCTGCTGCGTTCCATACACCTTTAGTTTCGTACGCGCAAAAACCATTTGCTAAAGCGG

2890 2900 2910 2920 2930 2940
TTGATAGCGCTAAATTTAAAGCGCCAAGCATTCCAGTGTTTGCTAATGGCACAGGCTTGG

2950 2960 2970 2980 2990 3000
TGCATTCAAGCAAACCGAATGACATTAAGAAAAACCTGAAAAACCACATGCTGGAATCTG

3010 3020 3030 3040 3050 3060
TTCATTTCAATCAAGAAATTGACAACATCTATGCTGATGGTGGCCGCGTATTTATCGAAT

3070 3080 3090 3100 3110 3120
TTGGTCCAAAGAATGTATTAATAAATTGGTTGAAAACATTCTCACTGAAAAATCTGATG

3130 3140 3150 3160 3170 3180
TGACTGCTATCGCGGTTAATGCTAATCCTAAACAACCTGCGGACGTACAAATGCGCCAAG

3190 3200 3210 3220 3230 3240
CTGCGCTGCAAATGGCAGTGCTTGGTGTCGCATTAGACAATATTGACCCGTACGACGCCG

3250 3260 3270 3280 3290 3300
TTAAGCGTCCACTTGTGCGCCGAAAGCATCACCAATGTTGATGAAGTTATCTGCAGCGT

3310 3320 3330 3340 3350 3360
CTTATGTTAGTCCGAAAACGAAGAAAGCGTTTGCTGATGCATTGACTGATGGCTGGACTG

3370 3380 3390 3400 3410 3420
TTAAGCAAGCGAAAGCTGTACCTGCTGTTGTGTACACCACAAGTGATTGAAAAGATCG

Fig. 6

3430 3440 3450 3460 3470 3480
TTGAAGTTGAAAAGATAGTTGAACGCATTGTCTGAAGTAGAGCGTATTGTCTGAAGTAGAAA

3490 3500 3510 3520 3530 3540
AAATCGTCTACGTTAATGCTGACGGTTCGCTTATATCGCAAAATAATCAAGACGTTAACA

3550 3560 3570 3580 3590 3600
GCGCTGTTGTTAGCAACGTGACTAATAGCTCAGTGACTCATAGCAGTGATGCTGACCTTG

3610 3620 3630 3640 3650 3660
TTGCCTCTATTGAACGCAGTGTTGGTCAATTTGTTGCACACCAACAGCAATTATTAAATG

3670 3680 3690 3700 3710 3720
TACATGAACAGTTTATGCAAGGTCCACAAGACTACGCGAAAACAGTGCAGAACGTACTTG

3730 3740 3750 3760 3770 3780
CTGCGCAGACGAGCAATGAATTACCGGAAAGTTTAGACCGTACATTGTCTATGTATAACG

3790 3800 3810 3820 3830 3840
AGTTCCAATCAGAAACGCTACGTGTACATGAAACGTACCTGAACAATCAGACGAGCAACA

3850 3860 3870 3880 3890 3900
TGAACACCATGCTTACTGGTGTGAAGCTGATGTGCTAGCAACCCCAATAACTCAGGTAG

3910 3920 3930 3940 3950 3960
TGAATACAGCCGTTGCCACTAGTCACAAGGTAGTTGCTCCAGTTATTGCTAATACAGTGA

3970 3980 3990 4000 4010 4020
CGAATGTTGTATCTAGTGTCTAGTAATAACGCGCGGTTGCAGTGCAAACCTGTGGCATTAG

4030 4040 4050 4060 4070 4080
CGCCTACGCAAGAAATCGCTCCAACAGTCGCTACTACGCCAGCACCCGCATTGGTTGCTA

4090 4100 4110 4120 4130 4140
TCGTGGCTGAACCTGTGATTGTTGCGCATGTTGCTACAGAAGTTGCACCAATTACACCAT

4150 4160 4170 4180 4190 4200
CAGTTACACCAGTTGTCTGCAACTCAAGCGGCTATCGATGTAGCAACTATTAACAAAGTAA

4210 4220 4230 4240 4250 4260
TGTTAGAAGTTGTTGCTGATAAAACCGGTTATCCAACGGATATGCTGGAACCTGAGCATGG

4270 4280 4290 4300 4310 4320
ACATGGAAGCTGACTTAGGTATCGACTCAATCAAACGTGTTGAGATATTAGGCGCAGTAC

4330 4340 4350 4360 4370 4380
AGGAATTGATCCCTGACTTACCTGAACCTTAATCCTGAAGATCTTGCTGAGCTACGCACGC

4390 4400 4410 4420 4430 4440
TTGGTGAGATTGTCGATTACATGAATTCAAAGCCCAGGCTGTAGCTCCTACAACAGTAC

4450 4460 4470 4480 4490 4500
CTGTAACAAGTGCACCTGTTTCGCCCTGCATCTGCTGGTATTGATTTAGCCCATCCAAA

4510 4520 4530 4540 4550 4560
ACGTAATGTTAGAAGTGGTTGCAGACAAAACGGTTACCCAACAGACATGCTAGAACTGA

Fig. 6

4570 4580 4590 4600 4610 4620
GCATGGATATGGAAGCTGACTTAGGTATTGATTCAATCAAGCGTGTGGAAATCTTAGGTG

4630 4640 4650 4660 4670 4680
CAGTACAGGAGATCATAACTGATTTACCTGAGCTAAACCCCTGAAGATCTTGCTGAATTAC

4690 4700 4710 4720 4730 4740
GCACCCTAGGTGAAATCGTTAGTTACATGCAAAGCAAAGCGCCAGTCGCTGAAAGTGCGC

4750 4760 4770 4780 4790 4800
CAGTGGCGACGGCTCCTGTAGCAACAAGCTCAGCACCGTCTATCGATTGAACCACATTC

4810 4820 4830 4840 4850 4860
AAACAGTGATGATGGATGTAGTTGCAGATAAGACTGGTTATCCAACTGACATGCTAGAAC

4870 4880 4890 4900 4910 4920
TTGGCATGGACATGGAAGCTGATTTAGGTATCGATTCAATCAAACGTGTGGAAATATTAG

4930 4940 4950 4960 4970 4980
GCGCAGTGCAGGAGATCATCACTGATTTACCTGAGCTAAACCCAGAAGACCTCGCTGAAT

4990 5000 5010 5020 5030 5040
TACGCACGCTAGGTGAAATCGTTAGTTACATGCAAAGCAAAGCGCCAGTCGCTGAGAGTG

5050 5060 5070 5080 5090 5100
CGCCAGTAGCGACGGCTTCTGTAGCAACAAGCTCTGCACCGTCTATCGATTTAAACCATA

5110 5120 5130 5140 5150 5160
TCCAAACAGTGATGATGGAAGTGGTTCAGACAAAACCGGTTATCCAGTAGACATGTTAG

5170 5180 5190 5200 5210 5220
AACTTGCTATGGACATGGAAGCTGACCTAGGTATCGATTCAATCAAGCGTGTAGAAATTT

5230 5240 5250 5260 5270 5280
TAGGTGCGGTACAGGAAATCATTACTGACTTACCTGAGCTTAACCCTGAAGATCTTGCTG

5290 5300 5310 5320 5330 5340
AACTACGTACATTAGGTGAAATCGTTAGTTACATGCAAAGCAAAGCGCCCGTAGCTGAAG

5350 5360 5370 5380 5390 5400
CGCCTGCAGTACCTGTTGCAGTAGAAAGTGCACCTACTAGTGTAACAAGCTCAGCACCGT

5410 5420 5430 5440 5450 5460
CTATCGATTTAGACCACATCCAAAATGTAATGATGGATGTTGTTGCTGATAAGACTGGTT

5470 5480 5490 5500 5510 5520
ATCCTGCCAATATGCTTGAATTAGCAATGGACATGGAAGCCGACCTTGGTATTGATTCAA

5530 5540 5550 5560 5570 5580
TCAAGCGTGTTGAAATTCTAGGCGCGGTACAGGAGATCATTACTGATTTACCTGAACATA

5590 5600 5610 5620 5630 5640
ACCCAGAAGACTTAGCTGAACTACGTACGTTAGAAGAAATTGTAACCTACATGCAAAGCA

5650 5660 5670 5680 5690 5700
AGGCGAGTGGTGTACTGTAAATGTAGTGGCTAGCCCTGAAAATAATGCTGTATCAGATG

Fig. 6

5710 5720 5730 5740 5750 5760
CATTATGCAAAGCAATGTGGCGACTATCACAGCGGCCGAGAACATAAGGCGGAATTTA

5770 5780 5790 5800 5810 5820
AACCGGCGCCGAGCGCAACCGTTGCTATCTCTCGTCTAAGCTCTATCAGTAAAATAAGCC

5830 5840 5850 5860 5870 5880
AAGATTGTAAAGGTGCTAACGCCTTAATCGTAGCTGATGGCACTGATAATGCTGTGTAC

5890 5900 5910 5920 5930 5940
TTGCAGACCACCTATTGCAAACCTGGCTGGAATGTAAGTGCATTGCAACCACTTGGGTAG

5950 5960 5970 5980 5990 6000
CTGTAACAACGACGAAAGCATTTAATAAGTCAGTGAACCTGGTGACTTTAAATGGCGTTG

6010 6020 6030 6040 6050 6060
ATGAAACTGAAATCAACAACATTATTACTGCTAACGCACAATTGGATGCAGTTATCTATC

6070 6080 6090 6100 6110 6120
TGCACGCAAGTAGCGAAATTAATGCTATCGAATACCCACAAGCATCTAAGCAAGGCCTGA

6130 6140 6150 6160 6170 6180
TGTTAGCCTTCTTATTAGCGAAATTGAGTAAAGTAACTCAAGCCGCTAAAGTGCCTGGCG

6190 6200 6210 6220 6230 6240
CCTTTATGATTGTTACTCAGCAGGGTGGTTCATTAGGTTTTGATGATATCGATTCTGCTA

6250 6260 6270 6280 6290 6300
CAAGTCATGATGTGAAAACAGACCTAGTACAAAGCGGCTTAAACGGTTTAGTTAAGACAC

6310 6320 6330 6340 6350 6360
TGTCTCACGAGTGGGATAACGTATTCTGTCTGCGGTTGATATTGCTTCGTCATTAAACGG

6370 6380 6390 6400 6410 6420
CTGAACAAGTTGCAAGCCTTGTTAGTGATGAACTACTTGATGCTAACACTGTATTAAACAG

6430 6440 6450 6460 6470 6480
AAGTGGGTTATCAACAAGCTGGTAAAGGCCTTGAACGTATCACGTAACTGGTGTGGCTA

6490 6500 6510 6520 6530 6540
CTGACAGCTATGCATTAAACAGCTGGCAATAACATCGATGCTAACTCGGTATTTTTAGTGA

6550 6560 6570 6580 6590 6600
GTGGTGGCGCAAAAGGTGTAAGTGCACATTGTGTTGCTCGTATAGCTAAAGAATATCAGT

6610 6620 6630 6640 6650 6660
CTAAGTTCATCTTATTGGGACGTTCAACGTTCTCAAGTGACGAACCGAGCTGGGCAAGTG

6670 6680 6690 6700 6710 6720
GTATTACTGATGAAGCGGCGTTAAAGAAAGCAGCGATGCAGTCTTTGATTACAGCAGGTG

6730 6740 6750 6760 6770 6780
ATAAACCAACACCCGTTAAGATCGTACAGCTAATCAAACCAATCCAAGCTAATCGTGAAA

6790 6800 6810 6820 6830 6840
TTGCGCAAACCTTGCTGCAATTACCGCTGCTGGTGCCAAGCTGAATATGTTTCTGCAG

Fig. 6

6850 6860 6870 6880 6890 6900
ATGTA ACTAATGCAGCAAGCGTACAAATGGCAGTCGCTCCAGCTATCGCTAAGTTCGGTG

6910 6920 6930 6940 6950 6960
CAATCACTGGCATCATTCATGGCGCGGGTGTGTTAGCTGACCAATTCATTGAGCAAAAAA

6970 6980 6990 7000 7010 7020
CACTGAGTGATTTTGAGTCTGTTTACAGCACTAAAATTGACGGTTTGTTATCGCTACTAT

7030 7040 7050 7060 7070 7080
CAGTCACTGAAGCAAGCAACATCAAGCAATTGGTATTGTTCTCGTCAGCGGCTGGTTTCT

7090 7100 7110 7120 7130 7140
ACGGTAACCCCGGCCAGTCTGATTACTCGATTGCCAATGAGATCTTAAATAAAACCGCAT

7150 7160 7170 7180 7190 7200
ACCGCTTTAAATCATTGCACCCACAAGCTCAAGTATTGAGCTTTAACTGGGGTCCCTTGGG

7210 7220 7230 7240 7250 7260
ACGGTGGCATGGTAACGCCTGAGCTTAAACGTATGTTTGACCAACGTGGTGTTTACATTA

7270 7280 7290 7300 7310 7320
TTCCACTTGATGCAGGTGCACAGTTATTGCTGAATGAACTAGCCGCTAATGATAACCGTT

7330 7340 7350 7360 7370 7380
GTCCACAAATCCTCGTGGGTAATGACTTATCTAAAGATGCTAGCTCTGATCAAAAGTCTG

7390 7400 7410 7420 7430 7440
ATGAAAAGAGTACTGCTGTAAAAAAGCCACAAGTTAGTCGTTTATCAGATGCTTTAGTAA

7450 7460 7470 7480 7490 7500
CTAAAAGTATCAAAGCGACTAACAGTAGCTCTTTATCAAACAAGACTAGTGCTTTATCAG

7510 7520 7530 7540 7550 7560
ACAGTAGTGCTTTTCAGGTTAACGAAAACCACTTTTTAGCTGACCACATGATCAAAGGCA

7570 7580 7590 7600 7610 7620
ATCAGGTATTACCAACGGTATGCGCGATTGCTTGGATGAGTGATGCAGCAAAAGCGACTT

7630 7640 7650 7660 7670 7680
ATAGTAACCGAGACTGTGCATTGAAGTATGTCCGTTTTCGAAGACTATAAATTGTTTAAAG

7690 7700 7710 7720 7730 7740
GTGTGTTTTTTGATGGCAATGAGGCGGCGGATTACCAAATCCAATTGTGCGCTGTGACAA

7750 7760 7770 7780 7790 7800
GGGCGTCAGAACAGGATTCTGAAGTCCGTATTGCCGCAAAGATCTTTAGCCTGAAAAGTG

7810 7820 7830 7840 7850 7860
ACGGTAAACCTGTGTTTCATTATGCAGCGACAATATTGTTAGCAACTCAGCCACTTAATG

7870 7880 7890 7900 7910 7920
CTGTGAAGGTAGAACTTCCGACATTGACAGAAAGTGTTGATAGCAACAATAAAGTAACTG

7930 7940 7950 7960 7970 7980
ATGAAGCACAAAGCGTTATACAGCAATGGCACCTTGTTCCACGGTGAAAGTCTGCAGGGCA

Fig. 6

7990 8000 8010 8020 8030 8040
TTAAGCAGATATTAAGTTGTGACGACAAGGGCCTGCTATTGGCTTGTCAGATAACCGATG

8050 8060 8070 8080 8090 8100
TTGCAACAGCTAAGCAGGGATCCTTCCCGTTAGCTGACAACAATATCTTTGCCAATGATT

8110 8120 8130 8140 8150 8160
TGGTTTATCAGGCTATGTTGGTCTGGGTGCGCAAACAATTTGGTTTAGGTAGCTTACCTT

8170 8180 8190 8200 8210 8220
CGGTGACAACGGCTTGGACTGTGTATCGTGAAGTGGTGTAGATGAAGTATTTTATCTGC

8230 8240 8250 8260 8270 8280
AACTTAATGTTGTTGAGCATGATCTATTGGGTTTACGCGGCAGTAAAGCCCGTTGTGATA

8290 8300 8310 8320 8330 8340
TTCAATTGATTGCTGCTGATATGCAATTACTTGCCGAAGTGAAATCAGCGCAAGTCAGTG

8350 8360 8370 8380 8390 8400
TCAGTGACATTTTGAACGATATGTCATGATCGAGTAAATAATAACGATAGGCGTCATGGT

8410 8420 8430 8440 8450 8460
GAGCATGGCGTCTGCTTTCTTCATTTTTTAAACATTAACAATATTAATAGCTAAACGCGGT

8470 8480 8490 8500 8510 8520
TGCTTTAAACCAAGTAAACAAGTGCTTTTtagctattactattccaaacaggatattaaag

8530 8540 8550 8560 8570 8580
AGAATATGACGGAATTAGCTGTTATTGGTATGGATGCTAAATTTAGCGGACAAGACAATA

8590 8600 8610 8620 8630 8640
TTGACCGTGTGGAACGCGCTTTCTATGAAGGTGCTTATGTAGGTAATGTTAGCCGCGTTA

8650 8660 8670 8680 8690 8700
GTACCGAATCTAATGTTATTAGCAATGGCGAAGAACAAGTTATTACTGCCATGACAGTTC

8710 8720 8730 8740 8750 8760
TTAACTCTGTCAGTCTACTAGCGCAAACGAATCAGTTAAATATAGCTGATATCGCGGTGT

8770 8780 8790 8800 8810 8820
TGCTGATTGCTGATGTAAAAAGTGCTGATGATCAGCTTGTAGTCCAAATTGCATCAGCAA

8830 8840 8850 8860 8870 8880
TTGAAAAACAGTGTCGAGTTGTGTTGTTATTGCTGATTTAGGCCAAGCATTAAATCAAG

8890 8900 8910 8920 8930 8940
TAGCTGATTTAGTTAATAACCAAGACTGTCTGTGGCTGTAATTGGCATGAATAACTCGG

8950 8960 8970 8980 8990 9000
TTAATTTATCTCGTCATGATCTTGAATCTGTAAGTGAACAATCAGCTTTGATGAAACCT

9010 9020 9030 9040 9050 9060
TCAATGGTTATAACAATGTAGCTGGGTTTCGCGAGTTTACTTATCGCTTCAACTGCGTTTG

9070 9080 9090 9100 9110 9120
CCAATGCTAAGCAATGTTATATATACGCCAACATTAAGGGCTTCGCTCAATCGGGCGTAA

Fig. 6

9130 9140 9150 9160 9170 9180
ATGCTCAATTTAACGTTGGAAACATTAGCGATACTGCAAAGACCGCATTGCAGCAAGCTA

9190 9200 9210 9220 9230 9240
GCATAACTGCAGAGCAGGTTGGTTTGTAGAAAGTGTGTCAGCAGTCGCTGATTTCGGCAATCG

9250 9260 9270 9280 9290 9300
CATTGTCTGAAAGCCAAGGTTTAATGTCTGCTTATCATCATACGCAAACCTTGCATACTG

9310 9320 9330 9340 9350 9360
CATTAAGCAGTCCCCGTAGTGTGACTGGTGAAGGCGGGTGTTCACAGGTCGCAGGTT

9370 9380 9390 9400 9410 9420
TATTGAAATGTGTAATTGGTTTACATCAACGTTATATTCCGGCGATTAAAGATTGGCAAC

9430 9440 9450 9460 9470 9480
AACCAGAGTGACAATCAAATGTCACGGTGGCGGAATTCACCATTCTATATGCCTGTAGATG

9490 9500 9510 9520 9530 9540
CTCGACCTTGGTTCCACATGCTGATGGCTCTGCACACATTGCCGCTTATAGTTGTGTGA

9550 9560 9570 9580 9590 9600
CTGCTGACAGCTATTGTTCATATTCTTTTACAAGAAAACGTCTTACAAGAACTTGTGTTTGA

9610 9620 9630 9640 9650 9660
AAGAAACAGTCTTGCAAGATAATGACTTAACTGAAAGCAAGCTTCAGACTCTTGAACAAA

9670 9680 9690 9700 9710 9720
ACAAATCCAGTAGCTGATCTGCGCACTAATGGTTACTTTGCATCGAGCGAGTTAGCATTA

9730 9740 9750 9760 9770 9780
TCATAGTACAAGGTAATGACGAAGCACAATTACGCTGTGAATTAGAACTATTACAGGGC

9790 9800 9810 9820 9830 9840
AGTTAAGTACTACTGGCATAAGTACTATCAGTATTAAACAGATCGCAGCAGACTGTTATG

9850 9860 9870 9880 9890 9900
CCCGTAATGATACTAACAAAGCCTATAGCGCAGTGCTTATTGCCGAGACTGCTGAAGAGT

9910 9920 9930 9940 9950 9960
TAAGCAAAGAAATAACCTTGGCGTTTGTCTGGTATCGCTAGCGTGTTTAAATGAAGATGCTA

9970 9980 9990 10000 10010 10020
AAGAATGGAAAACCCCGAAGGGCAGTTATTTTACCGCGCAGCCTGCAAATAAACAGGCTG

10030 10040 10050 10060 10070 10080
CTAACAGCACACAGAATGGTGTACCTTCATGTACCCAGGTATTGGTGCTACATATGTTG

10090 10100 10110 10120 10130 10140
GTTTAGGGCGTGATCTATTTTCATCTATTCCACAGATTTATCAGCCTGTAGCGGCTTTAG

10150 10160 10170 10180 10190 10200
CCGATGACATTGGCGAAAGTCTAAAAGATACTTTACTTAAATCCACGCAGTATTAGTCGTC

10210 10220 10230 10240 10250 10260
ATAGCTTTAAAGAACTCAAGCAGTTGGATCTGGACCTGCGCGGTAAGTTAGCCAATATCG

Fig. 6

10270 10280 10290 10300 10310 10320
CTGAAGCCGGTGTGGGTTTGGCTTGTGTGTTACCAAGGTATTTGAAGAAGTCTTTGCCG

10330 10340 10350 10360 10370 10380
TTAAAGCTGACTTTGCTACAGGTTATAGCATGGGTGAAGTAAGCATGTATGCAGCACTAG

10390 10400 10410 10420 10430 10440
GCTGCTGGCAGCAACCGGGATTGATGAGTGCTCGCCTTGCAACAATCGAATACCTTTAATC

10450 10460 10470 10480 10490 10500
ATCAACTTTGCGGCGAGTTAAGAACAACCTACGTCAGCATTGGGGCATGGATGATGTAGCTA

10510 10520 10530 10540 10550 10560
ACGGTACGTTGAGCAGATCTGGGAAACCTATACCATTAGGCAACGATTGAACAGGTGCG

10570 10580 10590 10600 10610 10620
AAATTGCCTCTGCAGATGAAGATCGTGTGTATTGCACCATTATCAATACACCTGATAGCT

10630 10640 10650 10660 10670 10680
TGTTGTAGCCGGTTATCCAGAAGCCTGTCAGCGAGTCATTAAGAATTTAGGTGTGCGTG

10690 10700 10710 10720 10730 10740
CAATGGCATTGAATATGGCGAACGCAATTCACAGCGCGCCAGCTTATGCCGAATACGATC

10750 10760 10770 10780 10790 10800
ATATGGTTGAGCTATACCATATGGATGTTACTCCACGTATTAATACCAAGATGTATTCAA

10810 10820 10830 10840 10850 10860
GCTCATGTTATTTACCGATTCCACAACGCAGCAAAGCGATTTCCACAGTATTGCTAAAT

10870 10880 10890 10900 10910 10920
GTTTGTGTGATGTGGTGGATTTCCACGTTTGGTTAATACCTTACATGACAAAGGTGCGC

10930 10940 10950 10960 10970 10980
GGGTATTCATTGAAATGGGTCCAGGTCGTTTCGTTATGTAGCTGGGTAGATAAGATCTTAG

10990 11000 11010 11020 11030 11040
TTAATGGCGATGGCGATAATAAAAAGCAAAGCCAACATGTATCTGTTCTGTGAATGCCA

11050 11060 11070 11080 11090 11100
AAGGCACCAAGTGATGAACTTACTTATATTTCGTGCGATTGCTAAGTTAATTAGTCATGGCG

11110 11120 11130 11140 11150 11160
TGAATTTGAATTTAGATAGCTTGTTTAAACGGGTCAATCCTGGTTAAAGCAGGCCATATAG

11170 11180 11190 11200 11210 11220
CAAACACGAACAAATAGTCAACATCGATATCTAGCGCTGGTGAGTTATACCTCATTAGTT

11230 11240 11250 11260 11270 11280
GAAATATGGATTTAAAGAGAGTAATTATGGAAAATATTGCAGTAGTAGGTATTGCTAATT

11290 11300 11310 11320 11330 11340
TGTTCCCGGGCTCACAAGCACCGGATCAATTTTGGCAGCAATTGCTTGAACAACAAGATT

11350 11360 11370 11380 11390 11400
GCCGCAGTAAGGCGACCGCTGTTCAAATGGGGGTTGATCCTGCTAAATATACCGCCAACA

Fig. 6

11410 11420 11430 11440 11450 11460
AAGGTGACACAGATAAATTTTACTGTGTGCACGGCGGTTACATCAGTGATTTCAATTTTG

11470 11480 11490 11500 11510 11520
ATGCTTCAGGTTATCAACTCGATAATGATTATTTAGCCGGTTTAGATGACCTTAATCAAT

11530 11540 11550 11560 11570 11580
GGGGGCTTTATGTTACGAAACAAGCCCTTACCGATGCGGGTTATTGGGGCAGTACTGCAC

11590 11600 11610 11620 11630 11640
TAGAAAACGTGGTGTGATTTTAGGTAATTTGTCATTCCCAACTAAATCATCTAATCAGC

11650 11660 11670 11680 11690 11700
TGTTTATGCCTTTGTATCATCAAGTTGTTGATAATGCCTTAAAGGCGGTATTACATCCTG

11710 11720 11730 11740 11750 11760
ATTTTCAATTAACGCATTACACAGCACCGAAAAAACACATGCTGACAATGCATTAGTAG

11770 11780 11790 11800 11810 11820
CAGGTTATCCAGCTGCATTGATCGCGCAAGCGGCGGTCTTGGTGGTTCACATTTTGCAC

11830 11840 11850 11860 11870 11880
TGGATGCGGCTTGTGCTTCATCTTGTTATAGCGTTAAGTTAGCGTGTGATTACCTGCATA

11890 11900 11910 11920 11930 11940
CGGGTAAAGCCAACATGATGCTTGCTGGTGCAGTATCTGCAGCAGATCCTATGTTTCGTAA

11950 11960 11970 11980 11990 12000
ATATGGGTTTCTCGATATTCCAAGCTTACCCAGCTAACAATGTACATGCCCGTTTGACC

12010 12020 12030 12040 12050 12060
AAAATTCACAAGGTCTATTTGCCGGTGAAGGCGCGGGCATGATGGTATTGAAACGTCAA

12070 12080 12090 12100 12110 12120
GTGATGCAGTACGTGATGGTATCATATTTACGCCATTATTAAAGGCGGCGCATTATCGA

12130 12140 12150 12160 12170 12180
ATGACGGTAAAGGCGAGTTTGTATTAAGCCGAACACCAAGGGCCAAGTATTAGTATATG

12190 12200 12210 12220 12230 12240
AACGTGCTTATGCCGATGCAGATGTTGACCCGAGTACAGTTGACTATATTGAATGTCATG

12250 12260 12270 12280 12290 12300
CAACGGGCACACCTAAGGGTGACAATGTTGAATTGCGTTCGATGGAAACCTTTTTTCAGTC

12310 12320 12330 12340 12350 12360
GCGTAAATAACAAACCATTACTGGGCTCGGTTAAATCTAACCTTGGTCATTTGTTAACTG

12370 12380 12390 12400 12410 12420
CCGCTGGTATGCCTGGCATGACCAAAGCTATGTTAGCGCTAGGTAAAGGTCTTATTCCTG

12430 12440 12450 12460 12470 12480
CAACGATTAACCTAAAGCAACCACTGCAATCTAAAAACGGTTACTTTACTGGCGAGCAAA

12490 12500 12510 12520 12530 12540
TGCCAACGACGACTGTGTCTTGGCCAACAACTCCGGGTGCCAAGGCAGATAAACCGCGTA

Fig. 6

12550 12560 12570 12580 12590 12600
CCGCAGGTGTGAGCGTATTTGGTTTTGGTGGCAGCAACGCCCATTGGTATTACAACAGC

12610 12620 12630 12640 12650 12660
CAACGCAAACACTCGAGACTAATTTTAGTGTTGCTAAACCACGTGAGCCTTTGGCTATTA

12670 12680 12690 12700 12710 12720
TTGGTATGGACAGCCATTTTGGTAGTGCCAGTAATTTAGCGCAGTTCAAAACCTTATTAA

12730 12740 12750 12760 12770 12780
ATAATAATCAAAATACCTTCCGTGAATTACCAGAACAACGCTGGAAAGGCATGGAAAGTA

12790 12800 12810 12820 12830 12840
ACGCTAACGTCATGCAGTCGTTACAATTACGCAAAGCGCCTAAAGGCAGTTACGTTGAAC

12850 12860 12870 12880 12890 12900
AGCTAGATATTGATTTCTTGCGTTTTAAAGTACCGCCTAATGAAAAAGATTGCTTGATCC

12910 12920 12930 12940 12950 12960
CGCAACAGTTAATGATGATGCAAGTGGCAGACAATGCTGCGAAAGACGGAGGTCTAGTTG

12970 12980 12990 13000 13010 13020
AAGGTCGTAATGTTGCGGTATTAGTAGCGATGGGCATGGAAGTGAATTACATCAGTATC

13030 13040 13050 13060 13070 13080
GTGGTCGCGTTAATCTAACCACCCAAATTGAAGACAGCTTATTACAGCAAGGTATTAAAC

13090 13100 13110 13120 13130 13140
TGACTGTTGAGCAACGTGAAGAACTGACCAATATTGCTAAAGACGGTGTTGCCTCGGCTG

13150 13160 13170 13180 13190 13200
CACAGCTAAATCAGTATACGAGTTTCATTGGTAATATTATGGCGTCACGTATTTGGGCGT

13210 13220 13230 13240 13250 13260
TATGGGATTTTCTGGTCCCTGCTATTACCGTATCGGCTGAAGAAAACCTCTGTTTATCGTT

13270 13280 13290 13300 13310 13320
GTGTTGAATTAGCTGAAAATCTATTTCAAACCAGTGATGTTGAAGCCGTTATTATTGCTG

13330 13340 13350 13360 13370 13380
CTGTTGATTTGCTGTTCAATTGAAAACATTACTTTACGTCAGCACTACGGTCCAGTTA

13390 13400 13410 13420 13430 13440
ATGAAAAGGGATCTGTAAGTGAATGTGGTCCGGTTAATGAAAGCAGTTCAGTAACCAACA

13450 13460 13470 13480 13490 13500
ATATTCTTGATCAGCAACAATGGCTGGTGGGTGAAGGCGCAGCGGCTATTGTCGTTAAAC

13510 13520 13530 13540 13550 13560
CGTCATCGCAAGTCACTGCTGAGCAAGTTTATGCGCGTATTGATGCGGTGAGTTTTGCC

13570 13580 13590 13600 13610 13620
CTGGTAGCAATGCGAAAGCAATTACGATTGCAGCGGATAAAGCATTAACTTGTCTGGTA

13630 13640 13650 13660 13670 13680
TCAGTGCTGCTGATGTAGCTAGTGTTGAAGCACATGCAAGTGGTTTTAGTGCCGAAAATA

Fig. 6

13690 13700 13710 13720 13730 13740
ATGCTGAAAAACCGCGTTACCGACTTTATACCCAAGCGCAAGTATCAGTTCGGTGAAAG

13750 13760 13770 13780 13790 13800
CCAATATTGGTCATACGTTTAATGCCTCGGGTATGGCGAGTATTATTAACGCGCTGC

13810 13820 13830 13840 13850 13860
TGTTAGATCAGAATACGAGTCAAGATCAGAAAAGCAAACATATTGCTATTAACGGTCTAG

13870 13880 13890 13900 13910 13920
GTCGTGATAACAGCTGCGCGCATCTTATCTTATCGAGTTCAGCGCAAGCGCATCAAGTTG

13930 13940 13950 13960 13970 13980
CACCAGCGCCTGTATCTGGTATGGCCAAGCAACGCCCACAGTTAGTTAAACCATCAAAC

13990 14000 14010 14020 14030 14040
TCGGTGGTCAGTTAATTAGCAACGCGATTGTTAACAGTGCGAGTTCATCTTTACACGCTA

14050 14060 14070 14080 14090 14100
TTAAAGCGCAGTTTGGCGGTAAGCACTTAACAAAGTTAACCAGCCAGTGATGATGGATA

14110 14120 14130 14140 14150 14160
ACCTGAAGCCCCAAGGTATTAGCGCTCATGCAACCAATGAGTATGTGGTGACTGGAGCTG

14170 14180 14190 14200 14210 14220
CTAACACTCAAGCTTCTAACATTCAAGCATCTCATGTTCAAGCGTCAAGTCATGCACAAG

14230 14240 14250 14260 14270 14280
AGATAGCACCAAACCAAGTTCAAAATATGCAAGCTACAGCAGCCGCTGTAAGTTCACCCC

14290 14300 14310 14320 14330 14340
TTTCTCAACATCAACACACAGCGCAGCCCGTAGCGGCACCGAGCGTTGTTGGAGTGACTG

14350 14360 14370 14380 14390 14400
TGAAACATAAAGCAAGTAACCAAATTCATCAGCAAGCGTCTACGCATAAAGCATTTTTAG

14410 14420 14430 14440 14450 14460
AAAGTCGTTTAGCTGCACAGAAAAACCTATCGCAACTTGTTGAATTGCAAACCAAGCTGT

14470 14480 14490 14500 14510 14520
CAATCCAAACTGGTAGTGACAATACATCTAACAATACTGCGTCAACAAGCAATACAGTGC

14530 14540 14550 14560 14570 14580
TAACAAATCCTGTATCAGCAACGCCATTAACTTGTGTCTAATGCGCCTGTAGTAGCGA

14590 14600 14610 14620 14630 14640
CAAACCTAACCAAGTACAGAAGCAAAAGCGCAAGCAGCTGCTACACAAGCTGGTTTTTCAGA

14650 14660 14670 14680 14690 14700
TAAAGGACCTGTTGGTTACAACCTATCCACCGCTGCAGTTAATTGAACGTTATAATAAAC

14710 14720 14730 14740 14750 14760
CAGAAAACGTGATTTACGATCAAGCTGATTTGGTTGAATTCGCTGAAGGTGATATTGGTA

14770 14780 14790 14800 14810 14820
AGGTATTTGGTGCTGAATACAATATTATTGATGGCTATTTCGCGTCGTGTACGTCTGCCAA

Fig. 6

14830 14840 14850 14860 14870 14880
CCTCAGATTACTTGTAGTAACACGTGTTACTGAACTTGATGCCAAGGTGCATGAATACA

14890 14900 14910 14920 14930 14940
AGAAATCATACATGTGTACTGAATATGATGTGCCTGTTGATGCACCGTTCTTAATTGATG

14950 14960 14970 14980 14990 15000
GTCAGATCCCTTGGTCTGTTGCCGTCGAATCAGGCCAGTGTGATTTGATGTTGATTTCAT

15010 15020 15030 15040 15050 15060
ATATCGGTATTGATTTCCAAGCGAAAGGCCGAACGTGTTTACCGTTTACTTGATTGTGAAT

15070 15080 15090 15100 15110 15120
TAACTTTCCTTGAAGAGATGGCTTTTGGTGGCGATACTTTACGTTACGAGATCCACATTG

15130 15140 15150 15160 15170 15180
ATTCGTATGCACGTAACGGCGAGCAATTATTATTCTTCTTCCATTACGATTGTTACGTAG

15190 15200 15210 15220 15230 15240
GGGATAAGAAGGTACTTATCATGCGTAATGGTTGTGCTGGTTTCTTTACTGACGAAGAAC

15250 15260 15270 15280 15290 15300
TTTCTGATGGTAAAGGCGTTATTCATAACGACAAAGACAAAGCTGAGTTTAGCAATGCTG

15310 15320 15330 15340 15350 15360
TTAAATCATCATTACGCCGTTATTACAACATAACCGTGGTCAATACGATTATAACGACA

15370 15380 15390 15400 15410 15420
TGATGAAGTTGGTTAATGGTGATGTTGCCAGTTGTTTTGGTCCGCAATATGATCAAGGTG

15430 15440 15450 15460 15470 15480
GCCGTAATCCATCATTGAAATTCTCGTCTGAGAAGTTCTTGATGATTGAACGTATTACCA

15490 15500 15510 15520 15530 15540
AGATAGACCCCAACCGGTGGTCATTGGGGACTAGGCCTGTTAGAAGGTCAGAAAGATTTAG

15550 15560 15570 15580 15590 15600
ACCTTGAGCATTGGTATTTCCCTTGTCACTTTAAAGGTGATCAAGTAATGGCTGGTTTCGT

15610 15620 15630 15640 15650 15660
TGATGTCGGAAGGTTGTGGCCAAATGGCGATGTTCTTCATGCTGTCTCTTGGTATGCATA

15670 15680 15690 15700 15710 15720
CCAATGTGAACAACGCTCGTTTCCAACCACTACCAGGTGAATCACAAACGGTACGTTGTC

15730 15740 15750 15760 15770 15780
GTGGGCAAGTACTGCCACAGCGCAATACCTTAACTTACCGTATGGAAGTTACTGCGATGG

15790 15800 15810 15820 15830 15840
GTATGCATCCACAGCCATTCATGAAAGCTAATATTGATATTTTGCTTGACGGTAAAGTGG

15850 15860 15870 15880 15890 15900
TTGTTGATTTCAAAACTTGAGCGTGATGATCAGCGAACAAGATGAGCATTACAGATTACC

15910 15920 15930 15940 15950 15960
CTGTAACACTGCCGAGTAATGTGGCGCTTAAAGCGATTACTGCACCTGTTGCGTCAGTAG

Fig. 6

15970 15980 15990 16000 16010 16020
CACCAGCATCTTCACCCGCTAACAGCGCGGATCTAGACGAACGTGGTGTGAACCGTTTA

16030 16040 16050 16060 16070 16080
AGTTTCCTGAACGTCCGTTAATGCGTGTTGAGTCAGACTTGTCTGCACCGAAAAGCAAAG

16090 16100 16110 16120 16130 16140
GTGTGACACCGATTAAAGCATTTTGAAGCGCCTGCTGTTGCTGGTCATCATAGAGTGCCTA

16150 16160 16170 16180 16190 16200
ACCAAGCACCGTTTACACCTTGGCATATGTTGAGTTTGCGACGGGTAATATTTCTAACT

16210 16220 16230 16240 16250 16260
GTTTCGGTCCTGATTTTGATGTTTATGAAGGTCGTATTCCACCTCGTACACCTTGTGGCG

16270 16280 16290 16300 16310 16320
ATTTACAAGTTGTTACTCAGGTTGTAGAAGTGCAGGGCGAACGTCTTGATCTTAAAAATC

16330 16340 16350 16360 16370 16380
CATCAAGCTGTGTAGCTGAATACTATGTACCGGAAGACGCTTGGTACTTTACTAAAAACA

16390 16400 16410 16420 16430 16440
GCCATGAAAACCTGGATGCCTTATTCATTAATCATGGAAATTGCATTGCAACCAAATGGCT

16450 16460 16470 16480 16490 16500
TTATTTCTGGTTACATGGGCACGACGCTTAAATACCCTGAAAAAGATCTGTTCTTCCGTA

16510 16520 16530 16540 16550 16560
ACCTTGATGGTAGCGGCACGTTATTAAAGCAGATTGATTTACGCGGCAAGACCATTTGTGA

16570 16580 16590 16600 16610 16620
ATAAATCAGTCTTGGTTAGTACGGCTATTGCTGGTGGCGCGATTATTCAAAGTTTCACGT

16630 16640 16650 16660 16670 16680
TTGATATGTCTGTAGATGGCGAGCTATTTTATACTGGTAAAGCTGTATTTGGTTACTTTA

16690 16700 16710 16720 16730 16740
GTGGTGAATCACTGACTAACCAACTGGGCATTGATAACGGTAAAACGACTAATGCGTGGT

16750 16760 16770 16780 16790 16800
TTGTTGATAACAATACCCCCGCAGCGAATATTGATGTGTTTGATTTAACTAATCAGTCAT

16810 16820 16830 16840 16850 16860
TGGCTCTGTATAAAGCGCCTGTGGATAAACCAGCATTATAAATTGGCTGGTGGTCAGATGA

16870 16880 16890 16900 16910 16920
ACTTTATCGATACAGTGTGAGTGGTTGAAGGCGGTGGTAAAGCGGGCGTGGCTTATGTTT

16930 16940 16950 16960 16970 16980
ATGGCGAACGTACGATTGATGCTGATGATTGGTTCTTCCGTTATCACTTCCACCAAGATC

16990 17000 17010 17020 17030 17040
CGGTGATGCCAGGTTCAATAGGTGTTGAAGCTATTATTGAGTTGATGCAGACCTATGCGC

17050 17060 17070 17080 17090 17100
TTAAAAATGATTTGGGTGGCAAGTTTGCTAACCACGTTTCATTGCGCCGATGACGCAAG

Fig. 6

17110 17120 17130 17140 17150 17160
TTGATTGGAAATACCGTGGGCAAATTACGCCGCTGAATAAACAGATGTCACCTGGACGTGC

17170 17180 17190 17200 17210 17220
ATATCACTGAGATCGTGAATGACGCTGGTGAAGTGCGAATCGTTGGTGATGCGAATCTGT

17230 17240 17250 17260 17270 17280
CTAAAGATGGTCTGCGTATTTATGAAGTTAAAAACATCGTTTTAAGTATTGTTGAAGCGT

17290 17300 17310 17320 17330 17340
AAAGGGTCAAGTGTAACGTGCTTAAGCGCCGCATTGGTTAAAGACGCTTTGCACGCCGTG

17350 17360 17370 17380 17390 17400
AATCCGTCATGGAGGCTTGGGGTTGGCATCCATGCCAACAACAGCAAGCTTACTTTAAT

17410 17420 17430 17440 17450 17460
CAATACGGCTTGGTGTCCATTTAGACGCCCTCGAACTTAGTAGTTAATAGACAAAATAATT

17470 17480 17490 17500 17510 17520
TAGCTGTGGAATGAATATAGTAAGTAATCATTCGGCAGCTACAAAAAGGAATTAAGAAT

17530 17540 17550 17560 17570 17580
GTCGAGTTTAGGTTTTAACAATAACAACGCAATTAAGTGGGCTTGGAAGTAGATCCAGC

17590 17600 17610 17620 17630 17640
GTCAGTTCATACACAAGATGCAGAAATTAAAGCAGCTTTAATGGATCTAACTAAACCTCT

17650 17660 17670 17680 17690 17700
CTATGTGGCGAATAATTCAGGCGTAAGTGGTATAGCTAATCATACTCAGTAGCAGGTGC

17710 17720 17730 17740 17750 17760
GATCAGCAATAACATCGATGTTGATGTATTGGCGTTTGCGCAAAAGTTAAACCCAGAAGA

17770 17780 17790 17800 17810 17820
TCTGGGTGATGATGCTTACAAGAAACAGCACGGCGTTAAATATGCTTATCATGGCGGTGC

17830 17840 17850 17860 17870 17880
GATGGCAAATGGTATTGCCTCGGTTGAATTGGTTGTTGCGTTAGGTAAAGCAGGGCTGTT

17890 17900 17910 17920 17930 17940
ATGTTCAATTTGGTGCTGCAGGTCTAGTGCCTGATGCGGTTGAAGATGCAATTCGTCGTAT

17950 17960 17970 17980 17990 18000
TCAAGCTGAATTACCAAATGGCCCTTATGCGGTTAACTTGATCCATGCACCAGCAGAAGA

18010 18020 18030 18040 18050 18060
AGCATTAGAGCGTGGCGCGGTTGAACGTTTCCTAAAACCTTGGCGTCAAGACGGTAGAGGC

18070 18080 18090 18100 18110 18120
TTCAGCTTACCTTGGTTTAACTGAACACATTGTTTGGTATCGTGCTGCTGGTCTAACTAA

18130 18140 18150 18160 18170 18180
AAACGCAGATGGCAGTGTTAATATCGGTAACAAGGTTATCGCTAAAGTATCGCGTACCGA

18190 18200 18210 18220 18230 18240
AGTTGGTCGCCGCTTTATGGAACCTGCACCGCAAAAATTACTGGATAAGTTATTAGAACA

Fig. 6

18250 18260 18270 18280 18290 18300
AAATAAGATCACCCCTGAACAAGCTGCTTTAGCGTTGCTTGTACCTATGGCTGATGATAT

18310 18320 18330 18340 18350 18360
TACTGGGGAAGCGGATTCTGGTGGTCATACAGATAACCGTCCGTTTTTAACATTATTACC

18370 18380 18390 18400 18410 18420
GACGATTATTGGTCTGCGTGATGAAGTGCAAGCGAAGTATAACTTCTCTCCTGCATTACG

18430 18440 18450 18460 18470 18480
TGTTGGTGTCTGGTGGTGGTATCGGAACGCCTGAAGCAGCACTCGCTGCATTTAACATGGG

18490 18500 18510 18520 18530 18540
CGCGGCTTATATCGTTCTGGGTTCTGTGAATCAGGCGTGTGTTGAAGCGGGTGCATCTGA

18550 18560 18570 18580 18590 18600
ATATACTCGTAAACTGTTATCGACAGTTGAAATGGCTGATGTGACTATGGCACCTGCTGC

18610 18620 18630 18640 18650 18660
AGATATGTTTGAAATGGGTGTGAAGCTGCAAGTATTAAAACGCGGTTCTATGTTTCGCGAT

18670 18680 18690 18700 18710 18720
GCGTGCGAAGAACTGTATGACTTGTATGTGGCTTATGACTCGATTGAAGATATCCCAGC

18730 18740 18750 18760 18770 18780
TGCTGAACGTGAGAAGATTGAAAAACAAATCTTCCGTGCAAACCTAGACGAGATTTGGGA

18790 18800 18810 18820 18830 18840
TGGCACTATCGCTTTCTTTACTGAACGCGATCCAGAAATGCTAGCCCGTGCAACGAGTAG

18850 18860 18870 18880 18890 18900
TCCTAAACGTAAATGGCACTTATCTTCCGTTGGTATCTTGGCCTTTCTTCACGCTGGTC

18910 18920 18930 18940 18950 18960
AAACACAGGCGAGAAGGGACGTGAAATGGATTATCAGATTTGGGCAGGCCCAAGTTTAGG

18970 18980 18990 19000 19010 19020
TGCATTCAACAGCTGGGTGAAAGGTTCTTACCTTGAAGACTATACCCGCCGTGGCGCTGT

19030 19040 19050 19060 19070 19080
AGATGTTGCTTTGCATATGCTTAAAGGTGCTGCGTATTTACAACGTGTAAACCAGTTGAA

19090 19100 19110 19120 19130 19140
ATTGCAAGGTGTTAGCTTAAGTACAGAATTGGCAAGTTATCGTACGAGTGATTAATGTTA

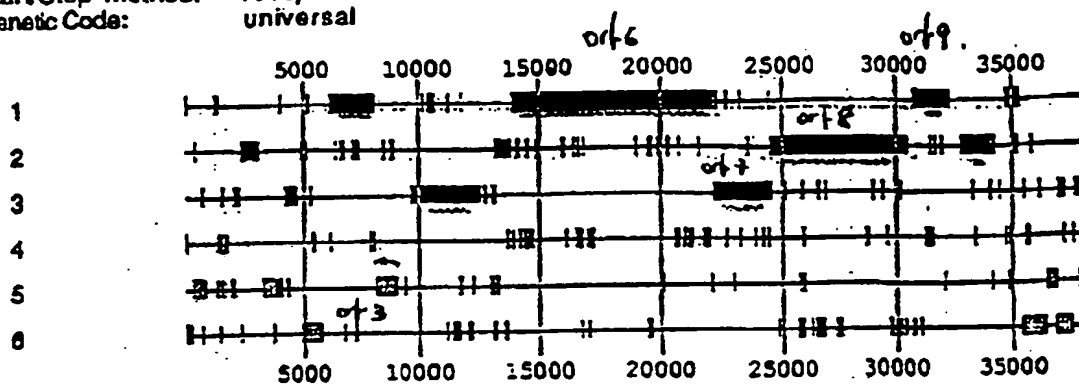
19150 19160 19170 19180 19190 19200
CTTGATGATATGTGAATTAATTAAAGCGCCTGAGGGCGCTTTTTTTGGTTTTTAACTCAG

19210 19220
GTGTTGTAACTCGAAATTGCCCCCTTC

Fig. 6

A

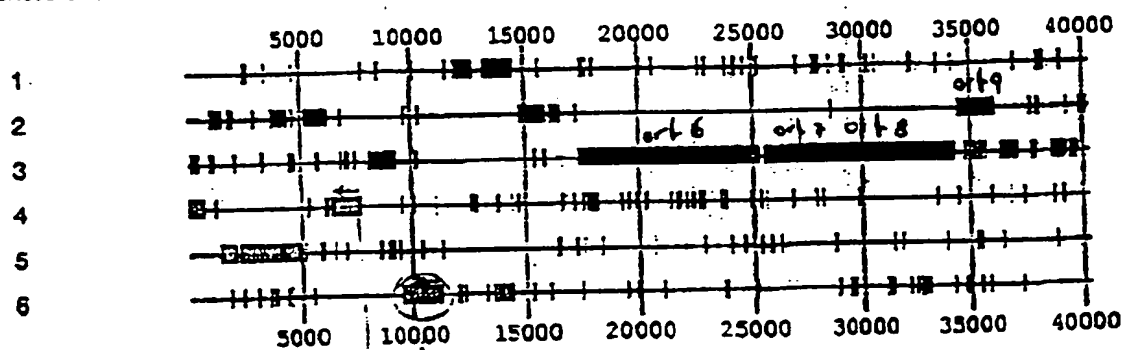
Start/Stop Method: AA span ≥ 25
Genetic Code: universal



Page 1

B

Start/Stop Method: AA span ≥ 25
Genetic Code: universal



A. A. A.

FIG 7

Start/Stop Method:
Genetic Code:

Avg. span ≥ 25
universal

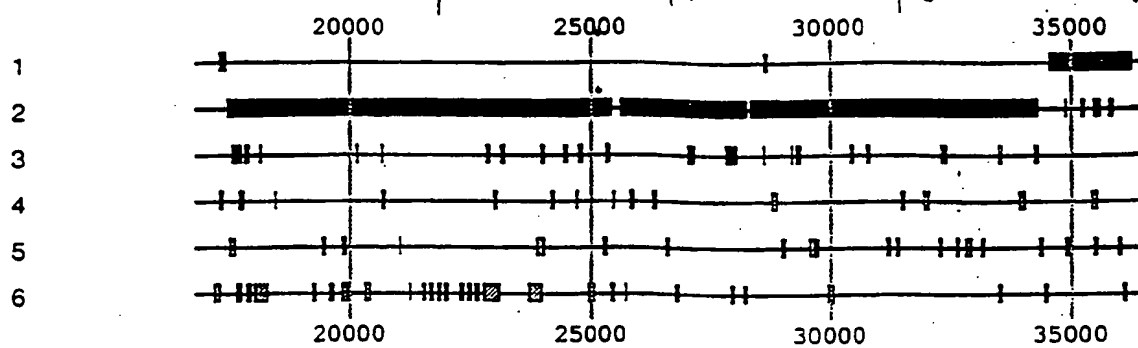


Fig. 8

Window Size = 8
Min. % Score = 60
Hash Value = 2

Scoring Matrix: BLOSUM 62

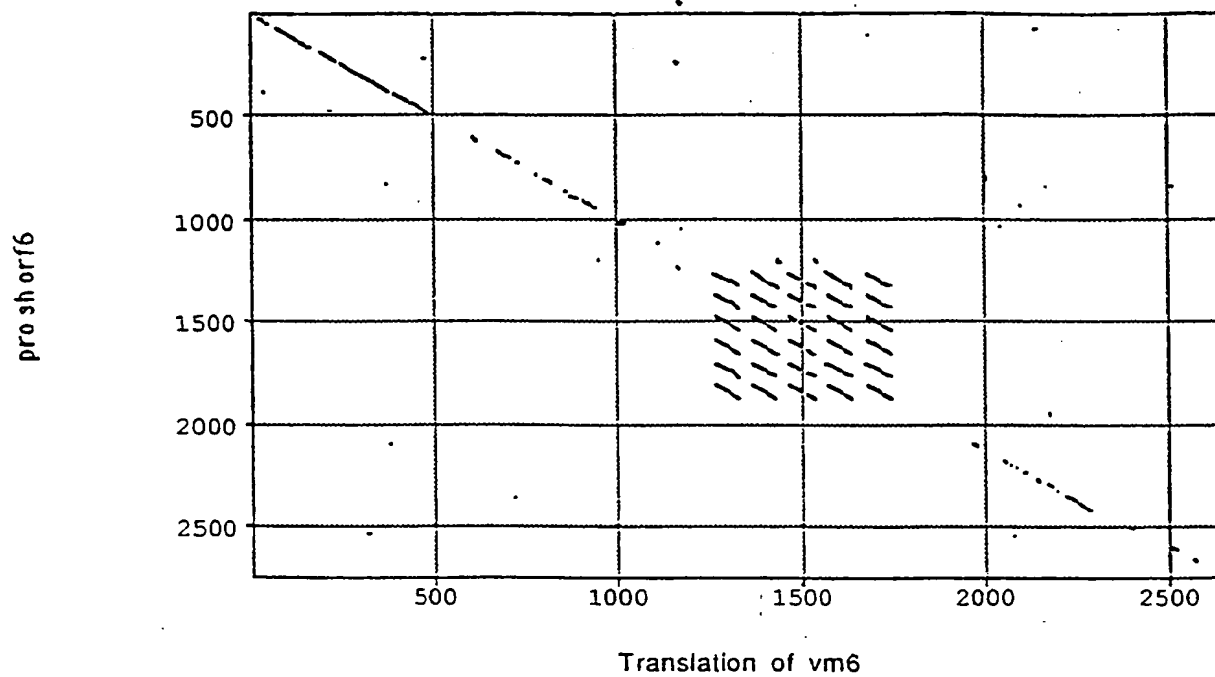


Fig. 9

Window Size = 8
Min. % Score = 60
Hash Value = 2

Scoring Matrix: BLOSUM 62

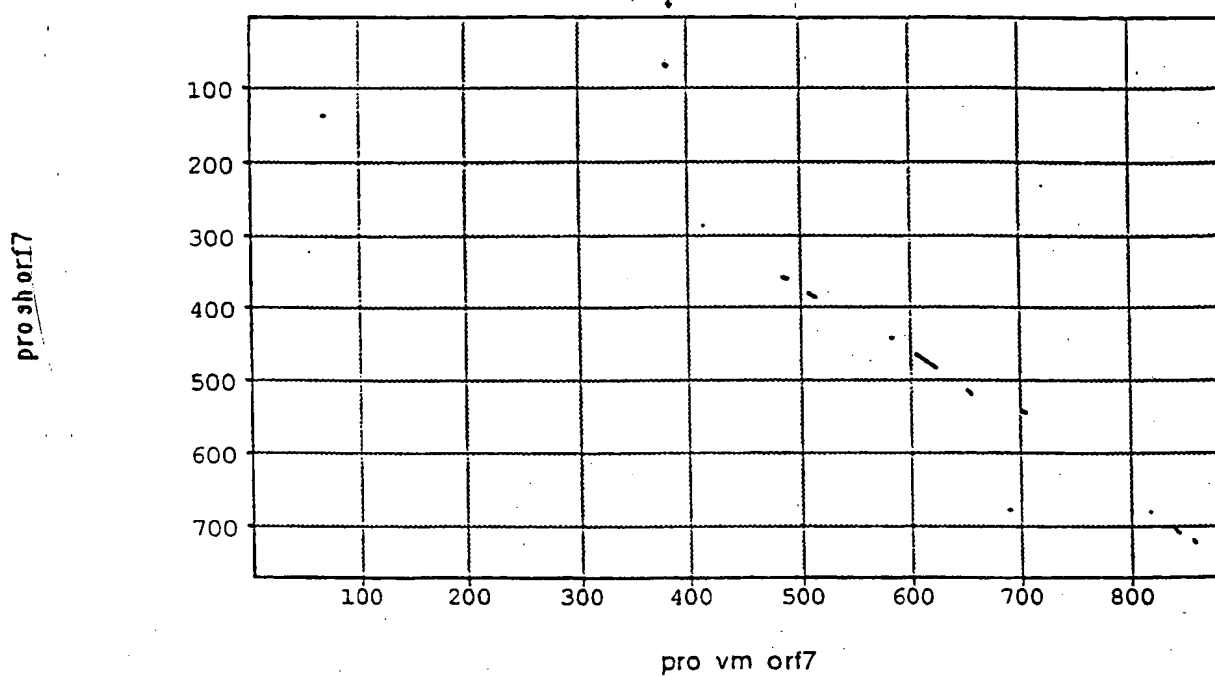


Fig. 10

Window Size = 8
Min. % Score = 60
Hash Value = 2

Scoring Matrix: BLOS 62

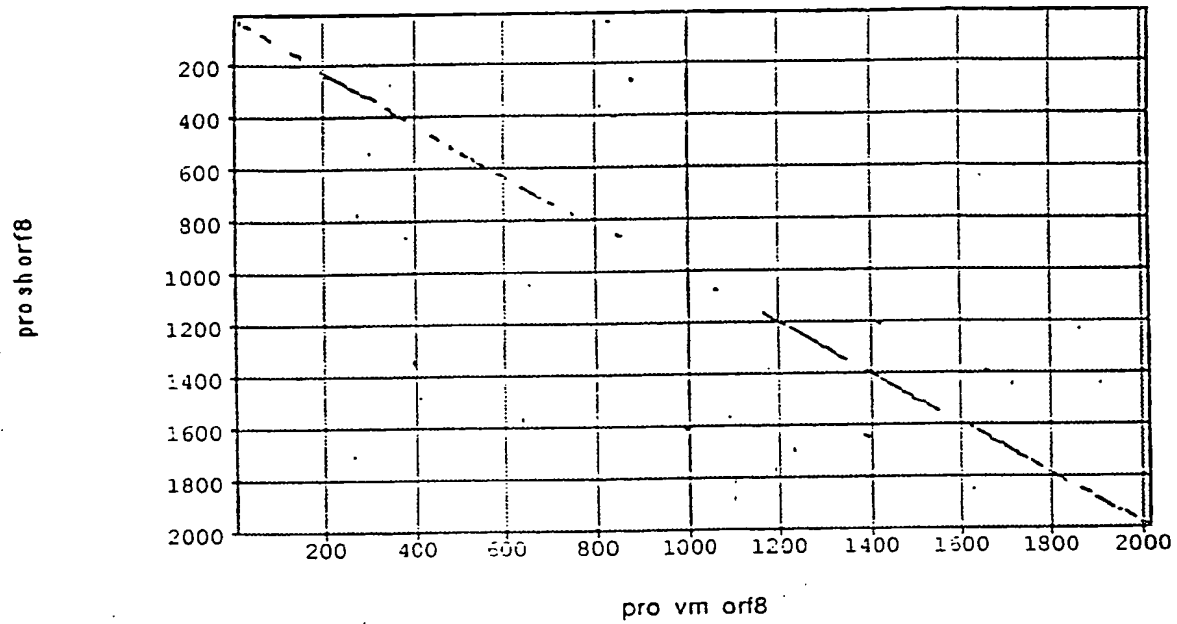


Fig. 11

Window Size = 8
Min. % Score = 60
Hash Value = 2

Scoring Matrix: BLO62

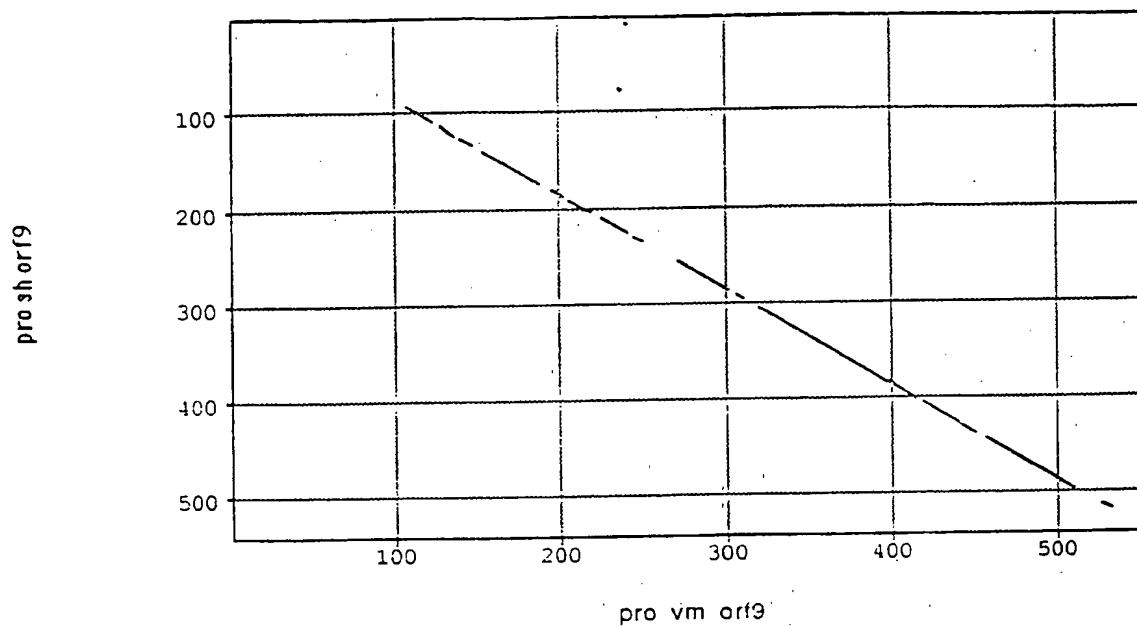


Fig. 12

COMPLEMENTATION Sp / Vm

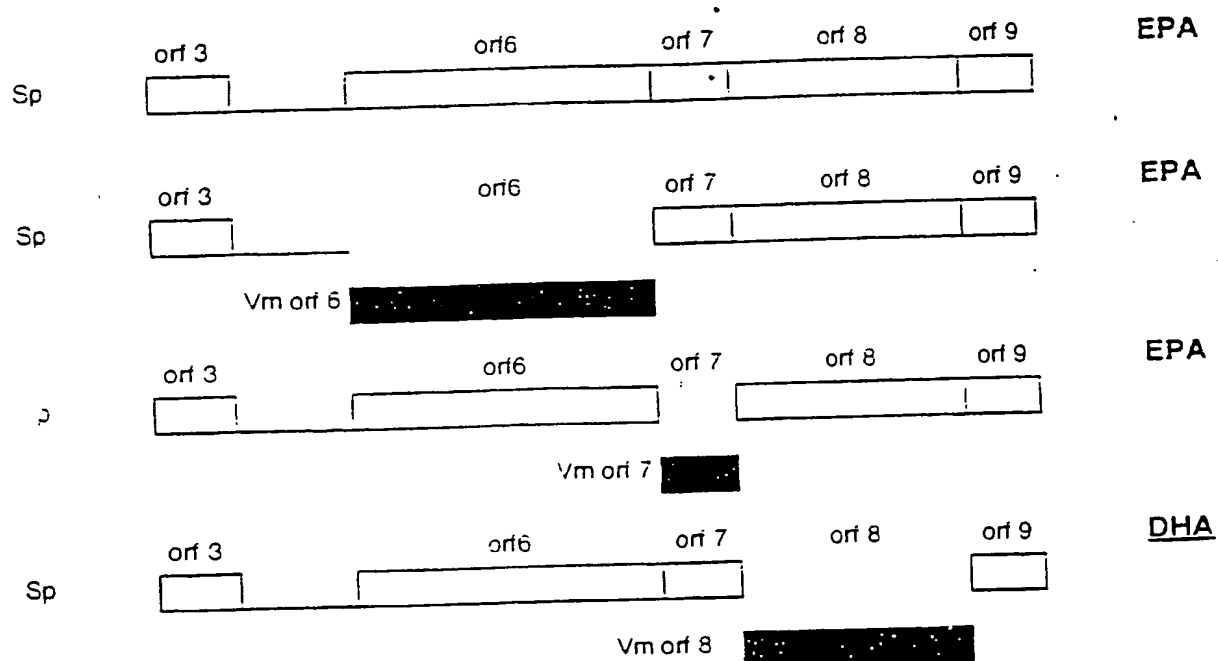


Fig. 13

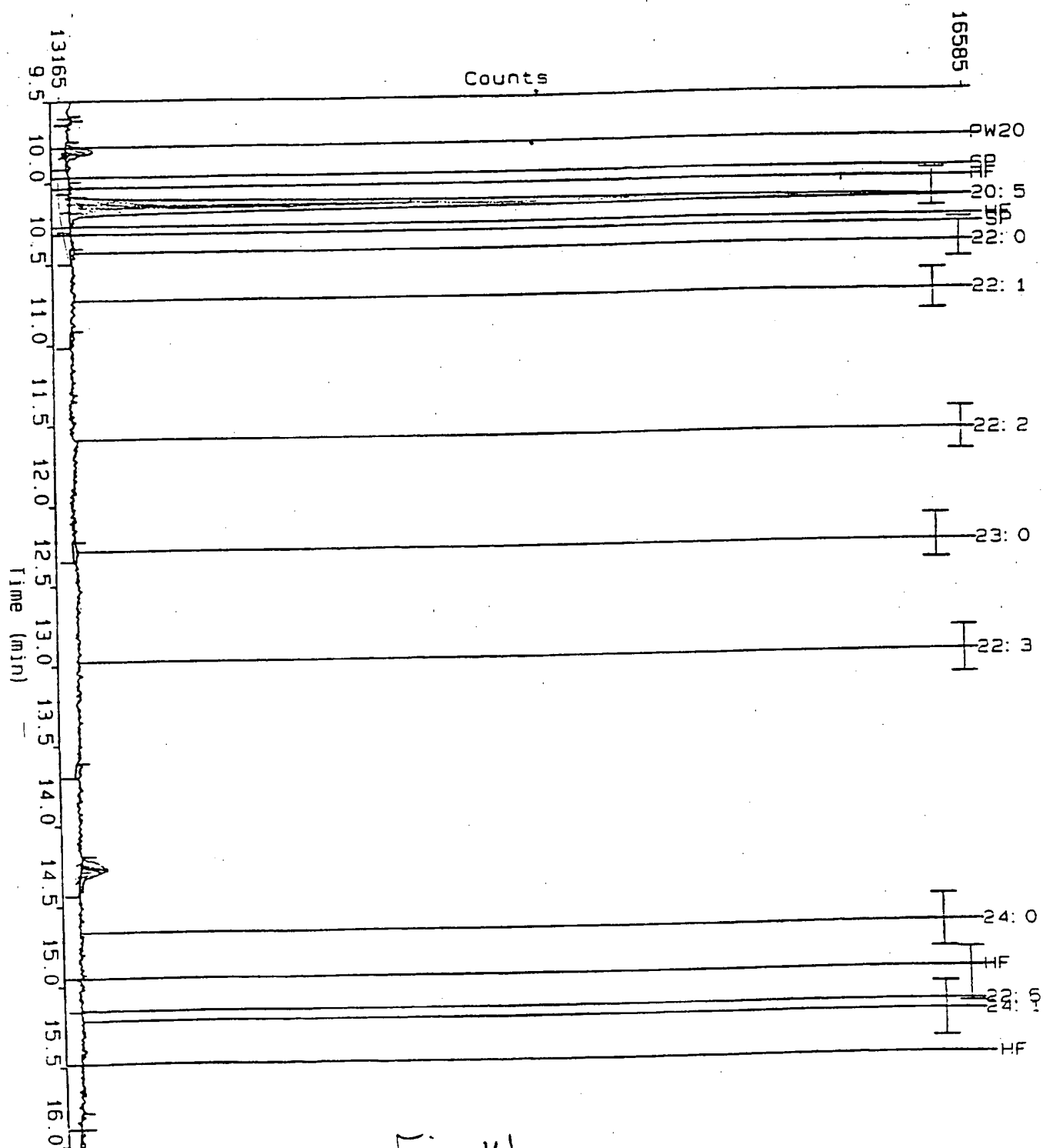


Fig. 14

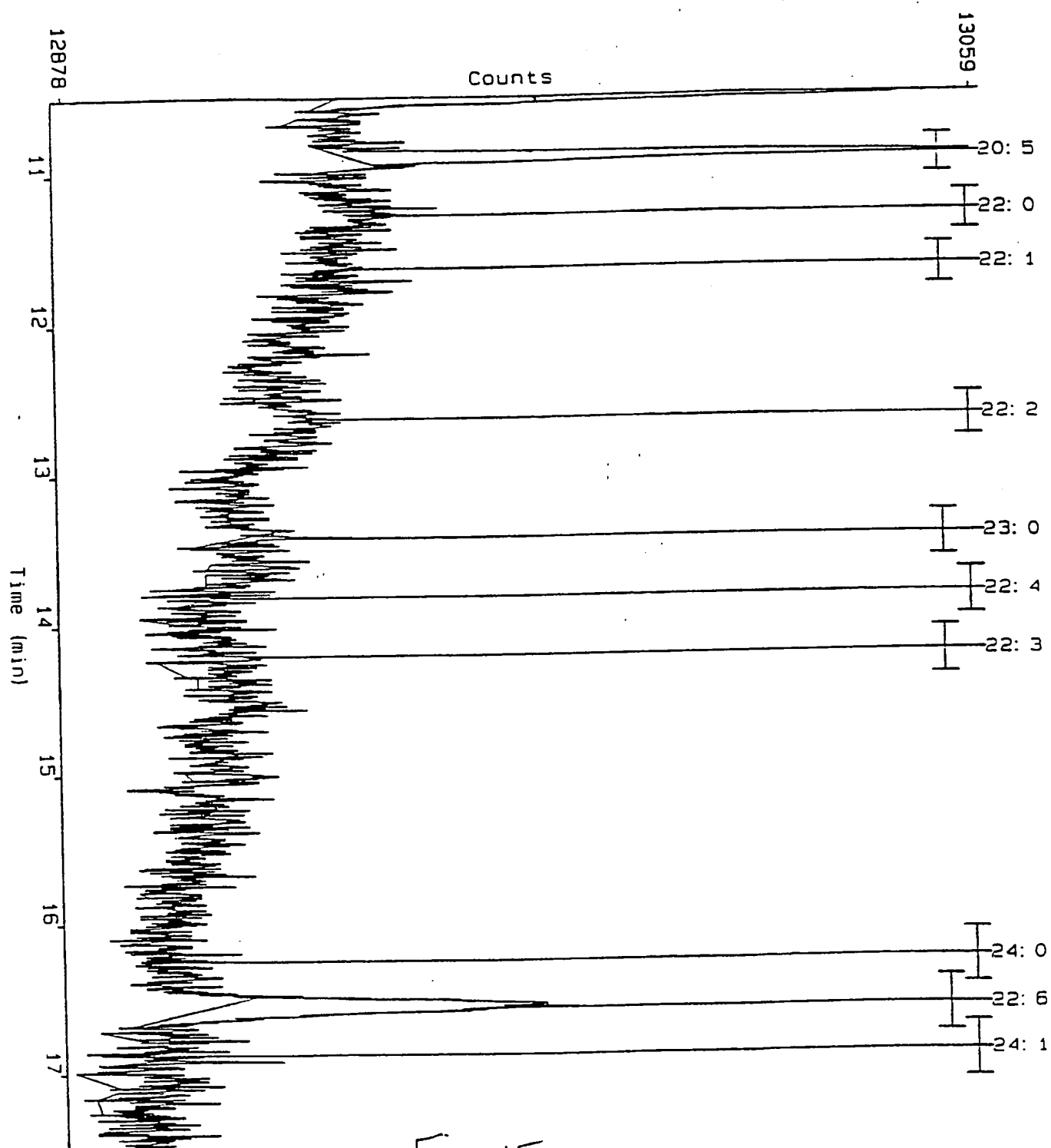


Fig. 15

<u>EPA (% Fatty acids)</u>	<u>DHA (% Fatty acids)</u>	<u>20°C</u>
0.00	0.06	pEPAD8
0.60	0.70	4
0.64	0.66	5
0.33	0.22	6s
0.45	0.59	6l
		<u>23°C</u>
0.02	0.06	pEPAD8
0.32	0.62	4
0.27	0.22	6s
0.18	0.65	6l

Fig. 16

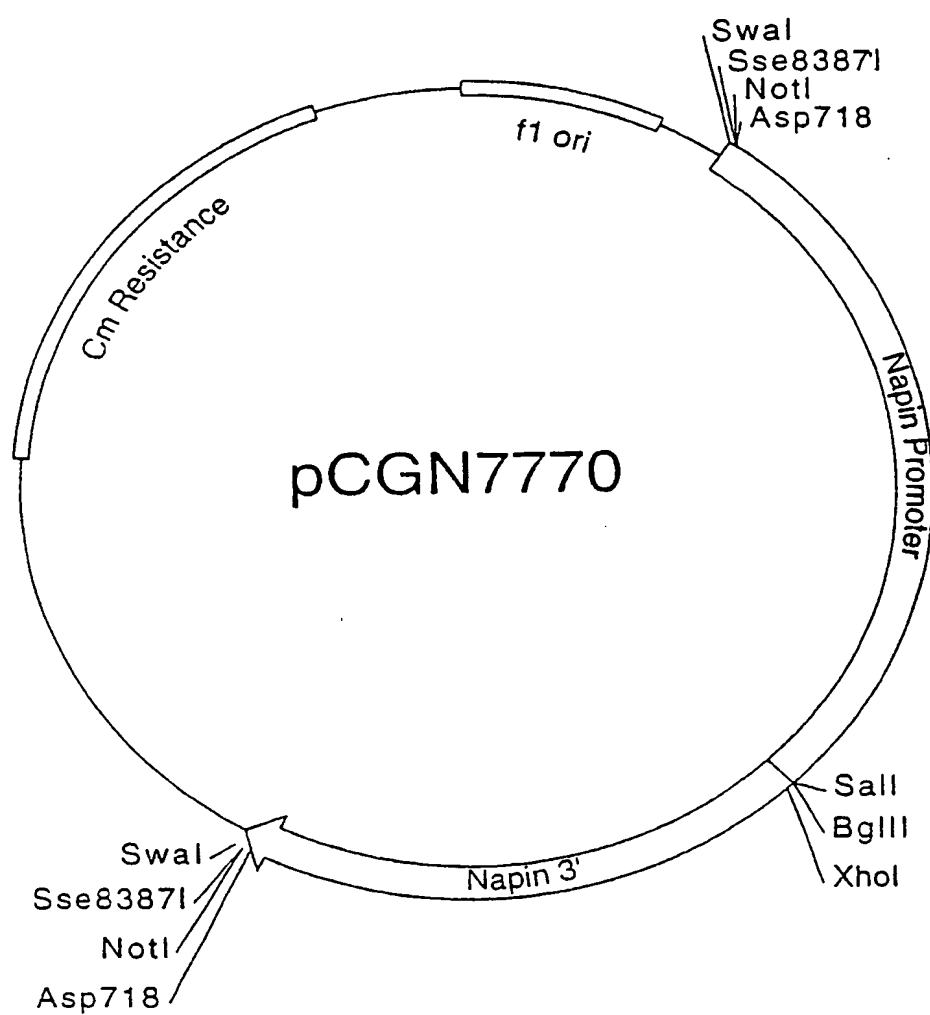
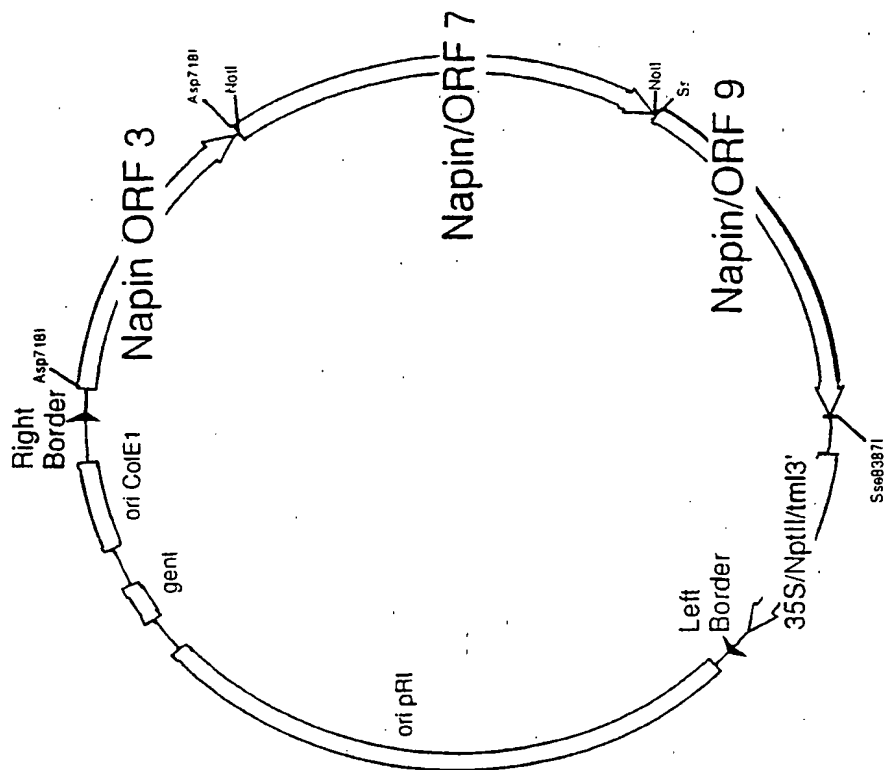
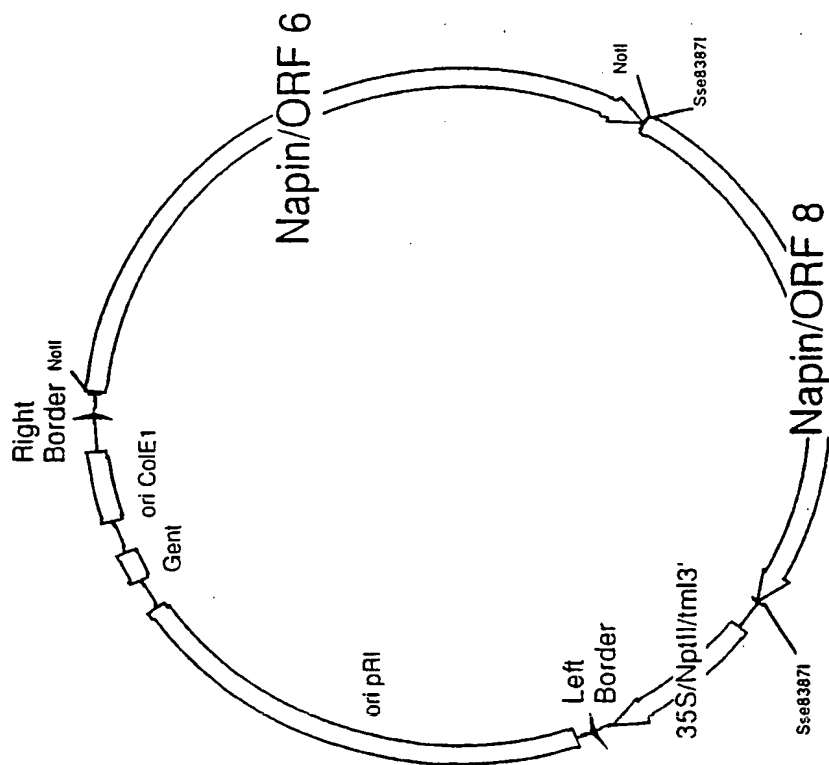


FIG 17

pCGN8535

pCGN8537



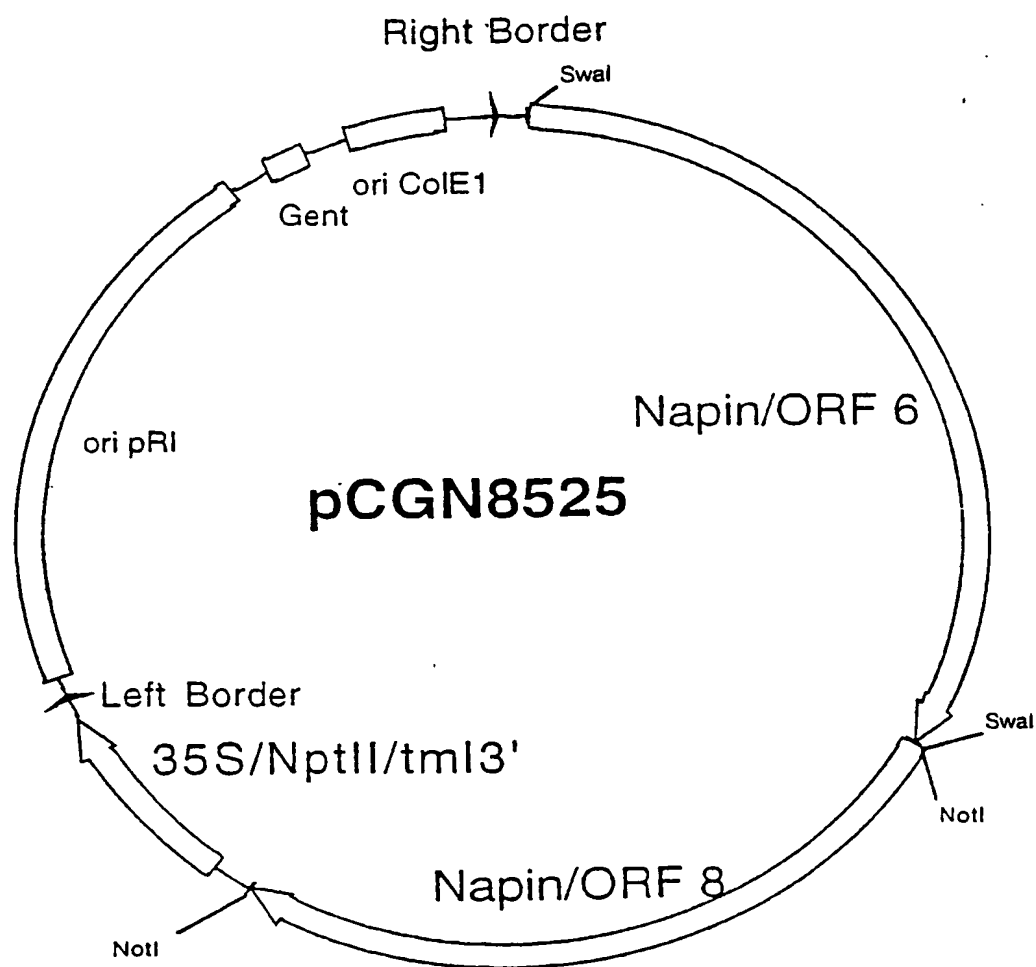


FIG 20

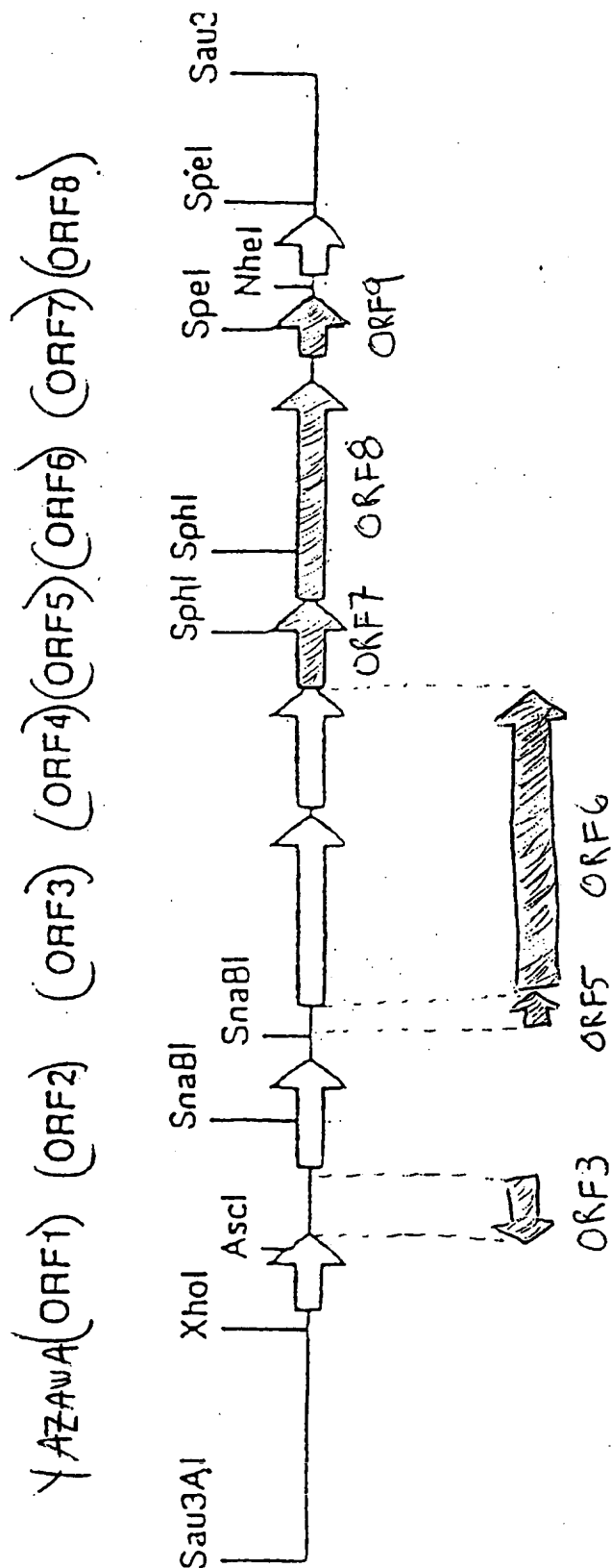


FIG 21

pCGN8556

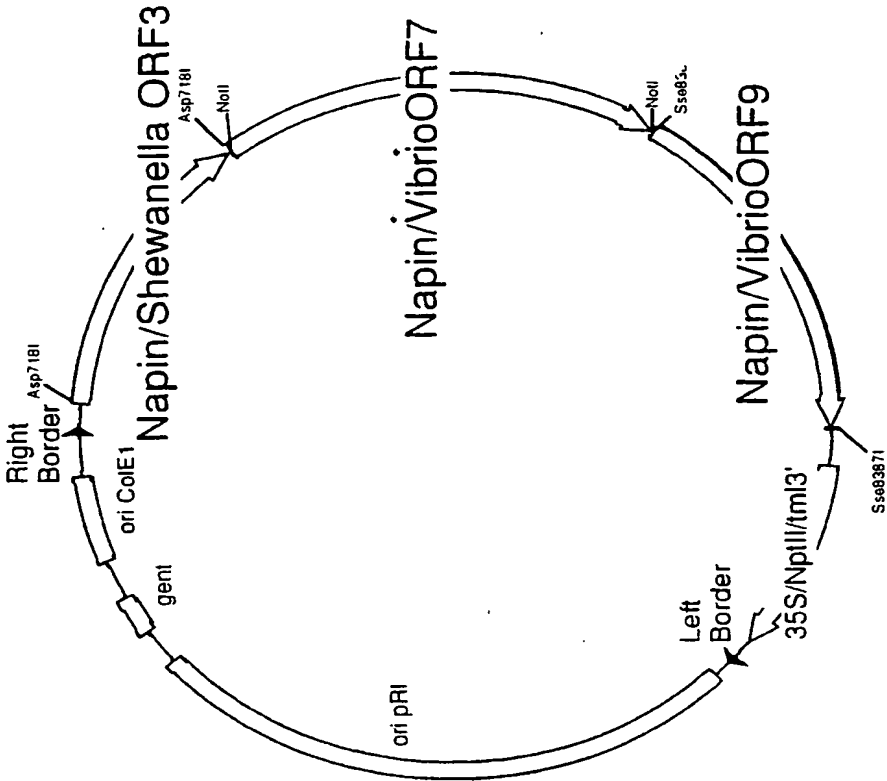


FIG 23

pCGN8560

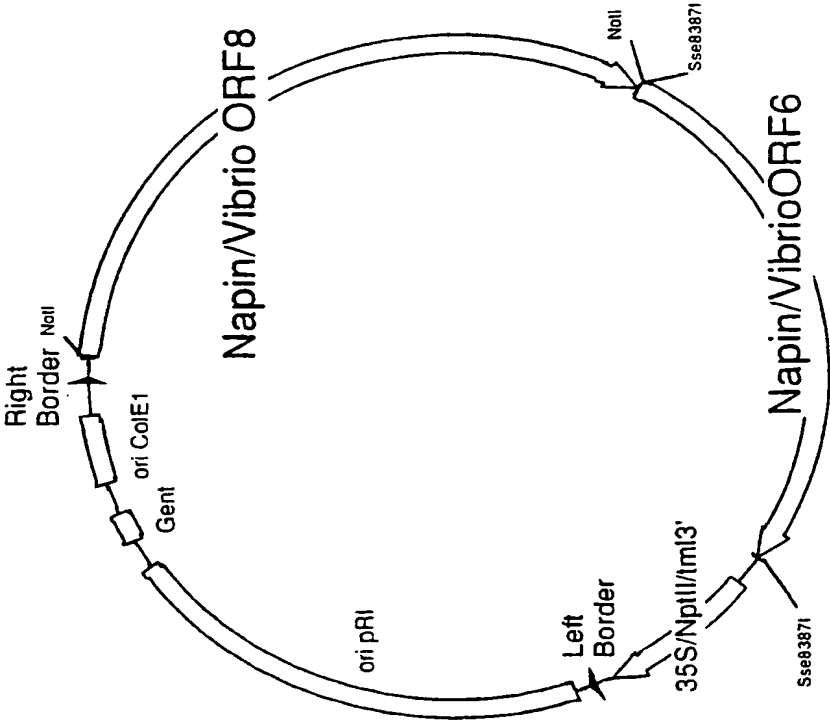


FIG 22

↓
ATT GGT AAA AAT AGG GGT TAT GTT TGT TGC TTT AAA GAG TGT CCT GAA
I G K N R G Y V C C F K E C P E>
↓
AAA TTG CTA ACT TCT CGA TTG ATT TCC TTA TAC TTC TGT CCG TTA ACA
K L L T S R L I S L Y F C P L T>
↓
ATA CAA GAG TGC GAT AAC CAG ACT ACA GAG TTG GTT AAG TCA TGG CTG
I Q E C D N Q T T E L V K S W L>
↓
CCT GAA GAT GAG TTA ATT AAG GTT AAT CGC TAC ATT AAA CAA GAA GCT
P E D E L I K V N R Y I K Q E A>
↓
9016
AAA ACT CAA GGT TTA ATG GTA AGA G
K T Q G L M V R>

FIG 24

10 20 30 40 50 60
AGCGAAATGCTTATCAAGAAATTCCAAGATCAATACATCACTGGGAAGAAAATTCATTCC
70 80 90 100 110 120
CTGGTTCACTGGGTAACGTTATTTCCGGCCGTATTGCTAACCGCTTCGACCTTGGTGGCA
130 140 150 160 170 180
TGAAGTGTGTCGTTGATGCAGCATGTGCAGGCCCTCTTGCTGCATTGCGTATGGCATTAA
190 200 210 220 230 240
GCGAGCTTGTTGAAGGCCCGCAGCGAAATGATGATTACAGGTGGTGTGTGTACCGATAACT
250 260 270 280 290 300
CACCAACCATGTACATGAGCTTCTCTAAACACCGGCATTACGACAAACGAAACAATTC
310 320 330 340 350 360
AACCATTGATATTGACTCGAAAGGTATGATGATTGGTGAAGGTATCGGTATGATTGCGC
370 380 390 400 410 420
TTAAACGTCTTGAAGACGCAGAGCGTGATGGCGAUCGTATCTATTCCGTGATTAAAGGTG
430 440 450 460 470 480
TTGGGTGCATCTTCAGACGGTAATTTATTAAGAGTANTTATGCGCNTCGTCTGAAGGTG
490 500 510 520 530 540
AGGCTAAGGCACTTAAACGTGCTTACGACGATGCAGGTTTCGCACCGCACACACTTGGCT
550 560 570 580 590 600
TACTTGAAGCCACGGCACAGGCACAGCAGGAGGTGATGTGGCAGAATTCAGTGGTCTTA
610 620 630 640 650 660
ACTCTGTATTCACTGAAGGCAATGACGAAAAGCAACACATCGCATTAGGTTCACTGAAAT
670 680 690 700 710 720
CACAGATTGGTCACTAAATCAACAGCGGGTACTGCGGGTCTAATCAAAGCGTCTTTAG
730 740 750 760 770 780
CACTGCACCATAAAGTACTGCCGCCAACCAATCAATGTAACCAGCCCTAACCTAAACTGA
790 800 810 820 830 840
ATATTGAAGACTCGCCTTTCTACCTCAATACACAGACGCGTCCATGGATGCAACGTGTGCG
850 860 870 880
ATGGTACACCGCGTCGTGCTGGTATTAGCTCATTGGTTTTGGTG

SS9 Photobacter

PCR Product Using Primers
Presented in Example I

FIG 25

3-21-VECTOR) by ORF. Ignored sequence
Tuesday, November 23, 1998 11:05 PM

Sequence Gaps: 1 to 408

```

      10      20      30      40      50      60
3-21-VECTOR  CCAAGTAAAG CCACTTAAGC CCACTTAAGC AGAGAGAGAG TTCTCTCTCT AAACATCTCT
      110      120 101 TO 808 OF ORF3-2
0. Jmp11832 CCAAGTAAAG CCACTTAAGC CCACTTAAGC AGAGAGAG TTCTCTCT AAACATCTCT
3-21-VECTOR  CCAAGTAAAG CCACTTAAGC CCACTTAAGC AGAGAGAG TTCTCTCT AAACATCTCT

      70      80      90      100     110     120
3-21-VECTOR  TCTAATTAAG CCACTTAAGC CCACTTAAGC AGAGAGAG TTCTCTCT AAACATCTCT
      170      180 101 TO 808 OF ORF3-2
0. Jmp11840 TCTAATTAAG CCACTTAAGC CCACTTAAGC AGAGAGAG TTCTCTCT AAACATCTCT
3-21-VECTOR  TCTAATTAAG CCACTTAAGC CCACTTAAGC AGAGAGAG TTCTCTCT AAACATCTCT
0. Jmp11848 TCTAATTAAG CCACTTAAGC CCACTTAAGC AGAGAGAG TTCTCTCT AAACATCTCT
3-21-VECTOR  TCTAATTAAG CCACTTAAGC CCACTTAAGC AGAGAGAG TTCTCTCT AAACATCTCT

      130     140     150     160     170     180
3-21-VECTOR  CCACTTAAGC CCACTTAAGC CCACTTAAGC AGAGAGAG TTCTCTCT AAACATCTCT
      230      240 101 TO 808 OF ORF3-2
0. Jmp11856 CCACTTAAGC CCACTTAAGC CCACTTAAGC AGAGAGAG TTCTCTCT AAACATCTCT
3-21-VECTOR  CCACTTAAGC CCACTTAAGC CCACTTAAGC AGAGAGAG TTCTCTCT AAACATCTCT
0. Jmp11864 CCACTTAAGC CCACTTAAGC CCACTTAAGC AGAGAGAG TTCTCTCT AAACATCTCT
3-21-VECTOR  CCACTTAAGC CCACTTAAGC CCACTTAAGC AGAGAGAG TTCTCTCT AAACATCTCT

      190     200     210     220     230     240
3-21-VECTOR  CCACTTAAGC CCACTTAAGC CCACTTAAGC AGAGAGAG TTCTCTCT AAACATCTCT
      290      300 101 TO 808 OF ORF3-2
0. Jmp11872 CCACTTAAGC CCACTTAAGC CCACTTAAGC AGAGAGAG TTCTCTCT AAACATCTCT
3-21-VECTOR  CCACTTAAGC CCACTTAAGC CCACTTAAGC AGAGAGAG TTCTCTCT AAACATCTCT
0. Jmp11880 CCACTTAAGC CCACTTAAGC CCACTTAAGC AGAGAGAG TTCTCTCT AAACATCTCT
3-21-VECTOR  CCACTTAAGC CCACTTAAGC CCACTTAAGC AGAGAGAG TTCTCTCT AAACATCTCT

      250     260     270     280     290     300
3-21-VECTOR  CCACTTAAGC CCACTTAAGC CCACTTAAGC AGAGAGAG TTCTCTCT AAACATCTCT
      350      360 101 TO 808 OF ORF3-2
0. Jmp11888 CCACTTAAGC CCACTTAAGC CCACTTAAGC AGAGAGAG TTCTCTCT AAACATCTCT
3-21-VECTOR  CCACTTAAGC CCACTTAAGC CCACTTAAGC AGAGAGAG TTCTCTCT AAACATCTCT
0. Jmp11896 CCACTTAAGC CCACTTAAGC CCACTTAAGC AGAGAGAG TTCTCTCT AAACATCTCT
3-21-VECTOR  CCACTTAAGC CCACTTAAGC CCACTTAAGC AGAGAGAG TTCTCTCT AAACATCTCT

      310     320     330     340     350     360
3-21-VECTOR  CCACTTAAGC CCACTTAAGC CCACTTAAGC AGAGAGAG TTCTCTCT AAACATCTCT
      410      420 101 TO 808 OF ORF3-2
0. Jmp11904 CCACTTAAGC CCACTTAAGC CCACTTAAGC AGAGAGAG TTCTCTCT AAACATCTCT
3-21-VECTOR  CCACTTAAGC CCACTTAAGC CCACTTAAGC AGAGAGAG TTCTCTCT AAACATCTCT
0. Jmp11912 CCACTTAAGC CCACTTAAGC CCACTTAAGC AGAGAGAG TTCTCTCT AAACATCTCT
3-21-VECTOR  CCACTTAAGC CCACTTAAGC CCACTTAAGC AGAGAGAG TTCTCTCT AAACATCTCT

```

ORF 6
Probe Resulting from PCR with Primers
Presented in Example I

FIG 26A

3-31-VECTO by GRP3 Allowed Sequence
 Tuesday, November 25, 7 11:08 AM

Page 3

```

0. jmp1 at      |      |      |      |      |
                | 1360 | 1365 | 1368 | 1370 |
-----|-----|-----|-----|-----|
3-31-VECTO      | 000000 00000000 00000000 00000000 00000000
                | 000000 00000000 00000000 00000000 00000000

          370      380      390      400
3-31-VECTO  00000000 00000000 00000000 00000000 00000000
          4700      101 TO 509 OF GRP3-3 5000
          |
0. jmp1380     | 1380 | 1400 | 1410 | 1420 |
-----|-----|-----|-----|
3-31-VECTO  00000000 00000000 00000000 00000000 00000000
          14000000 00000000 00000000 00000000 00000000

```

FIG 26B

INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 98/11639

A. CLASSIFICATION OF SUBJECT MATTER IPC 6 C12N15/31 C12N15/52 C12N15/82 C12N15/70 C12N5/10 C12N1/21 C12P7/64 A01H5/00		
According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) IPC 6 C12N C12P C07K A01H		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched		
Electronic data base consulted during the international search (name of data base and, where practical, search terms used)		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	NAKAHARA, TORO: "Physiological activity of docosahexaenoic acid (DHA) and its production by microbial culture" YUKAGAKU (1995), 44(10), 821-7 CODEN: YK GKAM; ISSN: 0513-398X, XP002080682 see abstract	6,7, 11-13
A	--- see abstract	14,32
X	NASU M ET AL: "Efficient transformation of Marchantia polymorpha that is haploid and has very small genome DNA; Agrobacterium tumefaciens-mediated transformation of suspension cell culture, for use in eicosapentaenoic acid, arachidonic acid and antibiotic production" J. FERMENT. BIOENG.; (1997) 84, 6, 519-23 CODEN: JFBIEX ISSN: 0922-338X, XP002080470 see the whole document --- -/--	25,27, 28,30
<input checked="" type="checkbox"/> Further documents are listed in the continuation of box C. <input checked="" type="checkbox"/> Patent family members are listed in annex.		
* Special categories of cited documents : "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier document but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art. "&" document member of the same patent family		
Date of the actual completion of the international search 14 October 1998		Date of mailing of the international search report 23/10/1998
Name and mailing address of the ISA European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016		Authorized officer Kania, T

INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 98/11639

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<p>KYLE D ET AL: "Long-chain omega-3 polyunsaturated fatty acids: prospects for introduction into horticultural food plants; e.g. alga eicosapentaenoic acid and docosahexaenoic acid gene cloning, expression in transgenic plant oil, crop improvement (conference paper)" HORTSCIENCE;(1990) 25, 12, 1523-26 CODEN: HJHSAR, XP002080471 * see the whole document, esp. p.1524, 2nd par. *</p>	25-28, 30,31
X	<p>EP 0 594 868 A (SAGAMI CHEM RES) 4 May 1994 cited in the application see the whole document</p>	15-17, 19-22,24
X	<p>WO 96 21735 A (SAGAMI CHEM RES) 18 July 1996 cited in the application see the whole document</p>	15-17, 19-22,24
A	<p>YAZAWA, KAZUNAGA: "Production of eicosapentaenoic acid from marine bacteria" LIPIDS (1996), 31(SUPPL., FATTY ACIDS AND LIPIDS FROM CELL BIOLOGY TO HUMAN DISEASE), S297-S300 CODEN: LPDSAP;ISSN: 0024-4201, XP002080483 cited in the application see the whole document</p>	1-32
A	<p>SOMERVILLE C R: "Future prospects for genetic modification of the composition of edible oils from higher plants; oilseed crop improvement by lipid and fatty acid modification (conference paper)" AM.J.CLIN.NUTR.:(1993) 58, 2, SUPPL., 270S-275S CODEN: AJCNAC, XP002080472 * see esp. p.274S, r. col., 1st par. *</p>	1-32

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/US 98/11639

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP 0594868 A	04-05-1994	AU 673359 B	07-11-1996
		AU 4088193 A	13-12-1993
		CA 2113557 A	25-11-1993
		FI 940203 A	14-03-1994
		WO 9323545 A	25-11-1993
		JP 6046864 A	22-02-1994
		NO 940146 A	14-03-1994
		US 5683898 A	04-11-1997
		US 5798259 A	25-08-1998
<hr/>			
WO 9621735 A	18-07-1996	AU 4400196 A	31-07-1996
		CA 2209987 A	18-07-1996
		EP 0831149 A	25-03-1998
		JP 8242867 A	24-09-1996
<hr/>			

THIS PAGE BLANK (USPTO)

This Page is inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

☒ BLACK BORDERS

☒ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES

☒ FADED TEXT OR DRAWING

☒ BLURED OR ILLEGIBLE TEXT OR DRAWING

☐ SKEWED/SLANTED IMAGES

☐ COLORED OR BLACK AND WHITE PHOTOGRAPHS

☐ GRAY SCALE DOCUMENTS

☐ LINES OR MARKS ON ORIGINAL DOCUMENT

☐ REPERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY

☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images
problems checked, please do not report the
problems to the IFW Image Problem Mailbox**

THIS PAGE BLANK (USPTO)